SD11
.A574
no.252

United States
Department of
Agriculture

Forest Service

General Technical
Report NC-252

March 2005

USDA

Proceedings of the Fourth Annual Forest Inventory and Analysis Symposium

Meeting jointly with the Southern Forest Mensurationists

New Orleans, Louisiana
November 19–21, 2002

United States
Department of
Agriculture

Forest Service

GTR–NC–252
February 2005

USDA

Proceedings of the Fourth Annual Forest Inventory and Analysis Symposium

Meeting jointly with the Southern Forest Mensurationists

Edited by
Ronald E. McRoberts
Gregory A. Reams
Paul C. Van Deusen
William H. McWilliams
Chris J. Cieszewski

New Orleans, Louisiana
November 19–21, 2002

ROCKY MOUNTAIN RESEARCH STATION LIBRARY
FORT COLLINS CENTER
240 W PROSPECT RD
FORT COLLINS, CO 80526-2098

Sponsored by
Forest Inventory and Analysis Program, USDA Forest Service
National Council for Air and Stream Improvement
Southern Forest Mensurationists
University of Georgia

McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H.; Cieszewski, Chris J., eds. 2005. Proceedings of the fourth annual forest inventory and analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 258 p.

Documents contributions to forest inventory in the areas of sampling, remote sensing, modeling, information management and analysis for the Forest Inventory and Analysis program of the USDA Forest Service.

KEY WORDS: annual forest inventory, estimation, sampling, modeling, remote sensing, assessments

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, sex, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326-W, Whitten Building, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.

The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

Papers published in these proceedings were submitted by authors in electronic media. Editing was done to ensure a consistent format. Authors are responsible for content and accuracy of their individual papers. The content does not necessarily reflect the views or policies of the U.S. Department of Agriculture or the Forest Service.

Preface

The Fourth Annual Forest Inventory and Analysis Symposium was held jointly with the meeting of the Southern Forest Mensurationists in New Orleans, Louisiana, November 19–21, 2002. The first three symposia of this series focused primarily on the statistical and remote sensing aspects of developing and implementing an annual forest inventory system. For this fourth symposium, a concerted effort was made to include presentations from resource analysts from within the Forest Inventory and Analysis (FIA) program and from researchers outside the FIA program who use FIA data. By all accounts, the fourth symposium achieved these objectives with numerous presentations that demonstrated the wide variety of applications for FIA data. The symposium organizers thank all presenters and convey special thanks to those who submitted their papers for this publication.

Ronald E. McRoberts
Gregory A. Reams
Paul C. Van Deusen
William H. McWilliams
Chris J. Cieszewski
St. Paul, Minnesota

Contents

Assessments

General Constraints on Sampling Wildlife on FIA Plots
Larissa L. Bailey, John R. Sauer, James D. Nichols, and Paul H. Geissler....1

The Application of FIA-based Data to Wildlife Habitat Modeling: A Comparative Study
Thomas C. Edwards, Jr.; Gretchen G. Moisen; Tracey S. Frescino; and Randall J. Schultz....7

Nonnative Plants in the Inventory of Western Oregon Forests
Andrew N. Gray....11

Providing Confidence in Regional Maps in Predicting Where Nonnative Species are Invading the Forested Landscape
Dennis M. Jacobs and Victor A. Rudis....17

Measuring Tree Seedlings and Associated Understory Vegetation in Pennsylvania's Forests
William H. McWilliams, Todd W. Bowersox, Patrick H. Brose, Daniel A. Devlin, James C. Finley, Kurt W. Gottschalk, Steve Horsley, Susan L. King, Brian M. LaPoint, Tonya W. Lister, Larry H. McCormick, Gary W. Miller, Charles T. Scott, Harry Steele, Kim C. Steiner, Susan L. Stout, James A. Westfall, and Robert L. White....21

Linking Soils and Down Woody Material Inventories for Cohesive Assessments of Ecosystem Carbon Pools
Katherine O'Neill, Christopher Woodall, Michael Amacher, and Geoffrey Holden....27

Clearcutting in the South: Issues, Status, and Trends
Jacek Siry and Frederick Cubbage....33

Modeling

The Effect of Data Quality on Short-term Growth Model Projections
David Gartner....41

Partitioning the Uncertainty in Estimates of Mean Basal Area Obtained from 10-year Diameter Growth Model Predictions
Ronald E. McRoberts....45

Individual-tree Green Weight Equations for Loblolly Pine (*Pinus Taeda* L.) Sawtimber in the Coastal Plain of Arkansas
Travis E. Posey, Paul F. Doruska, and David W. Patterson....53

Self-referencing Taper Curves for Loblolly Pine
Mike Strub, Chris Cieszewski, and David Hyink....59

D.B.H. and Survival Analysis: A New Methodology for Assessing Forest Inventory Mortality
Christopher W. Woodall, Patricia L. Grambsch, and William Thomas....65

Remote Sensing

Analysis of Pooled FIA and Remote Sensing Data for Fiber Supply Assessment at the Warnell School of Forest Resources at the University of Georgia—Other Studies and Effective Information Dissemination
Chris J. Cieszewski, Michal Zasada, Tripp Lowe, Bruce Borders, Mike Clutter, Richard F. Daniels, Robert I. Elle, Robert Izlar, and Jarek Zawadzki....71

Rapid Classification of Landsat TM Imagery for Phase 1 Stratification Using the Automated NDVI Threshold Supervised Classification (ANTSC) Methodology
William H. Cooke and Dennis M. Jacobs....81

Quantifying Forest Ground Flora Biomass Using Close-range Remote Sensing
Paul F. Doruska; Robert C. Weih, Jr.; Matthew D. Lane; and Don C. Bragg....87

Measuring Forest Area Loss Over Time Using FIA Plots and Satellite Imagery
Michael L. Hoppus and Andrew J. Lister....91

Spatially Locating FIA Plots from Pixel Values
Greg C. Liknes, Geoffrey R. Holden, Mark D. Nelson, and Ronald E. McRoberts....99

Variable Selection Strategies for Small-area Estimation Using FIA Plots and Remotely Sensed Data
Andrew Lister, Rachel Riemann, Jim Westfall, and Mike Hoppus 105

Distributing FIA Inventory Information onto Segmented Landsat Thematic Mapper Images Stratified with Industrial Ground Data
Tripp Lowe, Chris J. Cieszewski, Michal Zasada, and Jarek Zawadzki 111

Assessing the Effects of Forest Fragmentation Using Satellite Imagery and Forest Inventory Data
Ronald E. McRoberts and Greg C. Liknes 117

Comparing Forest/Nonforest Classifications of Landsat TM Imagery for Stratifying FIA Estimates of Forest Land Area
Mark D. Nelson, Ronald E. McRoberts, Greg C. Liknes, and Geoffrey R. Holden 121

Use of Semivariances for Studies of Landsat TM Image Textural Properties of Loblolly Pine Forests
Jarek Zawadzki, Chris J. Cieszewski, Roger C. Lowe, and Michal Zasada 129

Sampling and Estimation

Alternative Natural Resource Monitoring Strategies in the Mexican States of Jalisco and Colima
Cele Aguirre-Bravo and Hans Schreuder 135

Coordination, Cooperation, and Collaboration between FIA and NRI
Raymond L. Czaplewski, James Rack, Veronica C. Lessard, David F. Heinzen, Susan Ploetz, Thomas L. Schmidt, and Earl C. Leatherberry 141

The Effects of Removing Condition Boundaries on FIA Estimates
David Gartner and Gregory Reams 149

Comparison of Imputation Procedures for Replacing Denied-access Plots
Susan L. King 155

Strategies for Preserving Owner Privacy in the National Information Management System of the USDA Forest Service's Forest Inventory and Analysis Unit
Andrew Lister, Charles Scott, Susan King, Michael Hoppus, Brett Butler, and Douglas Griffith 163

COLE: A Web-based Tool for Interfacing with Forest Inventory Data
Patrick Proctor, Linda S. Heath, Paul C. Van Deusen, Jeffrey H. Gove, and James E. Smith 167

The Sensitivity of Derived Estimates to the Measurement Quality Objectives for Independent Variables
Francis A. Roesch 173

Modeling Missing Remeasurement Tree Heights in Forest Inventory Data
Raymond M. Sheffield and Callie J. Schweitzer 181

An Alternative View of Some FIA Sample Design and Analysis Issues
Paul C. Van Deusen 187

Sensitivity of FIA Volume Estimates to Changes in Stratum Weights and Number of Strata
James A. Westfall and Michael Hoppus 193

Sensitivity Analysis of Down Woody Material Data Processing Routines
Christopher W. Woodall and Duncan C. Lutes 199

Simulating Silvicultural Treatments Using FIA Data
Christopher W. Woodall and Carl E. Fiedler 203

Spatial Analyses

Spatial Information Needs on the Fishlake National Forest: Can FIA Help?
Robert B. Campbell, Jr., and Renee A. O'Brien 209

Predictive Mapping of Forest Attributes on the Fishlake National Forest
Tracey S. Frescino and Gretchen G. Moisen 215

TIGER 2000 and FIA
Joseph McCollum and Dennis Jacobs.....223

Regression and Geostatistical Techniques: Considerations and Observations from Experiences in NE-FIA
Rachel Riemann and Andrew Lister 231

Development and Validation of Spatially Explicit Habitat Models for Cavity-nesting Birds in Fishlake National Forest, Utah
Randall J. Schultz, Jr.; Thomas C. Edwards, Jr.; Gretchen G. Moisen; and Tracey S. Frescino 241

Integrating Spatial Components into FIA Models of Forest Resources: Some Technical Aspects
Pat Terletzky and Tracey Frescino 249

Using FIA and GIS Data to Estimate Areas and Volumes of Potential Stream Management Zones and Road Beautifying Buffers
Michal Zasada, Chris J. Cieszewski, Roger C. Lowe, Jarek Zawadzki, Mike Clutter, and Jacek P. Siry.....253

General Constraints on Sampling Wildlife on FIA Plots

Larissa L. Bailey[1], John R. Sauer[2], James D. Nichols[2], and Paul H. Geissler[3]

Abstract.—This paper reviews the constraints to sampling wildlife populations at FIA points. Wildlife sampling programs must have well-defined goals and provide information adequate to meet those goals. Investigators should choose a State variable based on information needs and the spatial sampling scale. We discuss estimation-based methods for three State variables: species richness, abundance, and patch occupancy. All methods incorporate two essential sources of variation: detectability estimation and spatial variation. FIA sampling imposes specific space and time criteria that may need to be adjusted to meet local wildlife objectives.

Traditionally, wildlife sampling programs have sought to document species distribution or abundance and monitor changes in those patterns over time. Management tended to focus on species thought to be declining (the "Declining Species Paradigm," Caughley 1994). Modern wildlife sampling programs focus on either management or scientific objectives (Yoccoz *et al.* 2001). Managers need survey data to understand population status, to evaluate the effects of current or past management actions, or to assist in predicting the consequences of proposed management actions. Scientific objectives generally focus on change in population in response to experimental manipulations or environmental change. The field of adaptive management combines science and management objectives by analyzing wildlife surveys to differentiate among competing scientific hypotheses of a system's response to management actions.

Recent criticism of wildlife sampling programs has focused on two main issues: 1) the lack of clearly defined goals; and 2) the need for estimation-based sampling methods (Olsen *et al.* 1999, Yoccoz *et al.* 2001). Researchers must consider their sampling objectives, and then choose a State variable to characterize the status of their biological system of interest. A State variable can be any variable within the system used to characterize and monitor the state of the system (e.g., population, species diversity, or biomass) (Yoccoz *et al.* 2001). Sampling methods must provide information adequate for estimating the chosen State variable. Traditional, entrenched sampling methods, such as point counts (birds) or time-constrained searches (amphibians and reptiles), are usually inappropriate for most sampling goals (Barker and Sauer 1995). While "standardization" of protocols is important to sampling, it does not ensure consistency in detection rates of most vertebrate species (Barker and Sauer 1995). Estimation-based methods that accommodate detectability differences and spatial variation are necessary to meet most wildlife sampling objectives.

The choice of an appropriate State variable depends on the sampling program's objectives, scale, and resources. Investigators must consider the biological level of the program's objectives and the spatial scale of the proposed sampling. For example, common State variables include species richness for community level analysis, abundance for population analysis, and patch occupancy rate for landscape or patch level analysis. Choosing a State variable also depends on the feasibility and efficiency of the sampling methods. Abundance is often the most expensive State variable to estimate in terms of time and effort and thus is rarely used in large programs. Occupancy and species richness estimation are less expensive and may be more appropriate for landscape-level studies. Occupancy estimation requires multiple visits to the same sites within a season, but only requires the collection of detection/nondetection information (MacKenzie *et al.* 2002). Species richness can also be estimated from detection/nondetection information from repeated visits (Boulinier *et al.* 1998) or using the counts of individuals of each species from one sample per location (Burnham and Overton 1979, Boulinier *et*

[1] Post-doctorate Researcher, USGS Cooperative Fish and Research Unit, North Carolina State University, Department of Zoology, Raleigh, NC 27695. Phone: 301–497–5637; e-mail: lbailey@usgs.gov.
[2] Wildlife Biologists, USGS Patuxent Wildlife Research Center, Laurel, MD 20708.
[3] Biometrician, USGS Patuxent Wildlife Research Center, Beltsville Laboratory-308, Beltsville, MD 20705.

al. 1998). However, species richness estimation is often ineffective in systems with small species pools.

Regardless of the State variable, sampling "protocols" should address two design components: the sample frame and detectability. The sample frame is a complete list of all possible sample units (i.e., plots, transects, quadrats, FIA points). Usually researchers cannot sample all possible units, so a subset of units is selected in some type of probabilistic manner (e.g., stratified random sample). Results from this subset of units are used to draw inferences about the entire area of interest. Likewise, survey methods rarely detect all the individuals or species in an area during the sampling interval. Detection probability, or "detectability," accounts for the proportion of animals or species missed during sampling. Historically, researchers have resisted estimating detection probability, claiming that it complicates field logistics and data analysis. Today, estimation methods are more accessible and new approaches are currently being developed. In the next section we detail current estimation-based methods for three common State variables.

Estimation-based Methods

Species Richness

Well-developed statistical theory exists for species richness estimation using capture-recapture methods (Boulinier *et al.* 1998, Williams *et al.* 2002). Each species is treated as an "individual" and either sample occasions (Boulinier *et al.* 1998) or spatial sample units (Nichols *et al.* 1998b) are treated as the "capture occasions." Models allow for detection probability to vary among sampling occasions (i.e., time), among species (heterogeneity), or by some "behavioral" response, or a combination of these factors (Boulinier *et al.* 1998). Behavioral response may occur when a species becomes more abundant or visible at some point during the sampling season, or if an observer's ability to detect a species increases with his/her exposure to that species (Boulinier *et al.* 1998). Sample units may be visited once or on multiple occasions. If units are visited only once, relative abundance data (counts of individuals for each species) can aid in species richness estimation (Boulinier *et al.* 1998, Burnham and Overton 1979). Theory exists for estimating spatial differences in species richness and community composition (Nichols *et al.* 1998b), and community dynamics such as extinction and colonization rates (Nichols *et al.* 1998a). Numerous computer programs for species richness estimation are available online at http://www.mbr-pwrc.usgs.gov/software.html. (See programs CAPTURE, SPECRICH, SPECRICH2, and COMDYN.)

Abundance Estimation

Extensive literature deals with abundance or population estimation (see Lancia *et al.* 1994, Nichols and Conroy 1996 for brief reviews). These methods can be categorized as either count-based or capture-recapture (or re-sighting) methods. Count-based estimation methods have seen a number of recent advances including distance sampling (Thomas *et al.* 2002), double-observer sampling (Nichols *et al.* 2000), temporal removal methods (Farnsworth *et al.* 2002), and replicated counts (Royle and Nichols 2003). Thompson (2002) reviews many of these methods and their application to sampling terrestrial bird populations.

Capture-recapture methods have existed for over two centuries (Seber 1982) and are typically classified as those suitable for closed (population constant during sampling) or open (population varying among sample periods) populations. Pollock *et al.* (1990), Pollock (2000), and Buckland *et al.* (2000) briefly review many of the capture-recapture models, including key references. There are many options for abundance estimation, and investigators should pay close attention to model assumptions and how they apply to their own systems. Combining estimation methods can increase model flexibility and help resolve problems with restrictive assumptions (Alpizar-Jara and Pollock 1996, Farnsworth *et al.* 2002, Powell *et al.* 2000).

There is an increasing variety of software availabe for analyzing abundance data including programs MARK (White and Burnham 1999, http://www.cnr.colostate.edu/~gwhite/mark/mark.html), DISTANCE (Buckland *et al.* 1993, http://www.ruwpa.stand.ac.uk/distance/), and others (see http://www.mbr-pwrc.usgs.gov/software.html). Most of these software packages have good technical assistance and online user's manuals.

Occupancy Estimation

For many species it is difficult to estimate abundance, but it is often feasible to estimate the proportion of the sample area where a species occurs. MacKenzie *et al.* (2002) developed a statistical-based method for estimating patch occupancy rates when species detection probabilities are less than 1. The method utilized detection/non-detection data from multiple visits to sites within a sampling season. Parameters include species-specific detection probabilities and the proportion of sites occupied. Sites can be discrete sampling units (ponds or patches of forest) or plots or quadrants chosen from an area of interest (MacKenzie *et al.* 2002). Occupancy rates may be modeled as a function of site-specific (habitat) covariates, while detection probabilities can be modeled as functions of both site-specific and time-specific covariates (e.g., sampling occasion, temperature, weather conditions) (MacKenzie *et al.* 2002). Theory also exists to estimate extinction and colonization rates, and occupancy change over time (MacKenzie *et al.* 2003). Current areas of research include estimating species co-occurrence rates, which could prove valuable in multi-species monitoring programs such as the Multiple-Species Inventory and Monitoring (MSIM) approach described in other papers from this forum. Program PRESENCE is specialized software designed for estimating and modeling occupancy rates using detection/nondetection information. It is available at: http://www.proteus.co.nz, and versions of the model have also been incorporated into program MARK (http://www.cnr.colostate.edu/~gwhite/mark/mark.html).

Traditional Sampling Example: Limitations and Modifications

Count-based sampling methods are widespread and deeply entrenched in wildlife literature. We use "point counts" as an example of a count-based technique to demonstrate its uses and limitations. We discuss ancillary information that can be collected to estimate detection probability.

Point counts are a traditional bird survey method used in a wide variety of studies including the North American Breeding Bird Survey (BBS). In point counts, a single observer counts the individuals detected (seen or heard) of each species within a fixed time period and sampling radius. Time periods and sample radii often vary among studies.

Species richness and occupancy rates can be estimated using point counts, especially if sites are visited several times within a sampling season; however, point counts are not appropriate for abundance estimation without estimating detectability. Varying traditional point count protocol would allow for detectability estimation including collecting distance information, using two observers, or recording time of detection. Variable circular points (VCP) are modified point counts where the distance to each detected bird is recorded. Detection rate is modeled as a function of distance from the point and used to adjust raw counts and estimate abundance. Unless all birds near the point are detected, a "donut effect" could result, causing biased abundance estimates. Biologists often question the accuracy of distance methods, but measurement error can be reduced by carefully training observers or using rangefinders. Program DISTANCE is a powerful software package designed to aid in analyzing distance sampling data and it can now incorporate GIS information (see http://www.ruwpa.stand.ac.uk/distance/).

Another way to estimate detection probabilities is to use the double-observer approach (Nichols *et al.* 2000). At each point count, a "primary" observer indicates to a "secondary" observer all birds detected. The secondary observer records all the primary observer's detections as well as any birds missed by the primary observer. Observers alternate primary and secondary roles. This approach permits estimating observer-specific detection probability rates that can be used to adjust raw counts and estimate abundance (Nichols *et al.* 2000).

Finally, a time removal method is a good option when most detections are by sound or when singing frequency is believed to be a major factor influencing detectability (Farnsworth *et al.* 2002). The timed point count is divided into intervals (equally spaced, if possible) where all individual bird detections are recorded in each interval. The method can estimate detection probabilities that vary by species, over time, or among observers (Farnsworth *et al.* 2002).

In summary, traditional, nonreplicated point counts are appropriate only for estimating species richness. If points are visited on multiple occasions within season, then both occupancy rate and species richness can be estimated. Further ancillary

information is necessary to estimate abundance, but investigators have a variety of options that could be tailored to their specific system of interest.

Application to Forest Inventory and Analysis Plots

The Forest Inventory and Analysis (FIA) program is useful because of its well-defined survey units and large-scale, long-term monitoring history. The ability to coordinate wildlife sampling at FIA points facilitates collection of co-located and historical information that may be important in determining: 1) species distributional range and temporal trends; 2) species-specific habitat associations; and 3) species response to management action. Unfortunately, FIA imposes a sample frame at a specific spatial and temporal scale. This scale is likely insufficient (i.e., too coarse) to meet local wildlife management or scientific objectives. Defining the scope of inference at FIA points in terms of wildlife populations is difficult. One alternative is to define a two-stage sampling procedure in which a group of FIA points constitutes a "local sampling frame" and draw samples from within the area (e.g., National Forests or ecoregions). This is the format used in the MSIM approach discussed in detail in other papers (e.g., Dunk et. al 2004, Manley et. al 2004). MSIM uses estimation-based methods to monitor occupancy rate as the focal State variable, but species richness could also be estimated under the current design. Hypotheses about spatial and temporal variations in these State variables can be tested statistically.

It is useful to consider alternative designs that provide more detailed spatial information at the scale of forests for use in local management. For example, a stratified frame could be developed with forest-scale monitoring of a chosen State variable in which FIA points were treated as a stratum within the forest. Again, the appropriate sampling frame will be influenced by the study's management or scientific objectives, and any sampling "protocol" applied to sites within the frame should incorporate estimation-based methods.

Literature Cited

Alpizar-Jara, Russell; Pollock, Kenneth H. 1996. A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys. Journal of Environmental and Ecological Statistics. 3(4): 311–327.

Barker, Richard J.; Sauer, John R. 1995. Statistical aspects of point count sampling. In: Ralph, C.J.; Sauer, J.R.; Droege, S., eds. Monitoring bird populations by point counts. Gen. Tech. Rep. PSW-149. Berkeley, CA: U. S. Department of Agriculture, Forest Service, Pacific Southwest Research Station: 125–130.

Boulinier, Thierry; Nichols, James D.; Sauer, John R.; et al. 1998. Estimating species richness: the importance of heterogeneity in species detectability. Ecology. 79(3): 1018–1028.

Buckland, Stephen T.; Goudie, Ian B.J.; Borchers, David L. 2000. Wildlife population assessment: past developments and future directions. Biometrics. 56(1): 1–12.

Burnham, Kenneth. P.; Overton W. S. 1979. Robust estimation of population-size when capture probabilities vary among animals. Ecology. 60(5): 927–936.

Caughley, Graeme. 1994. Directions in conservation biology. Journal of Animal Ecology. 63(2): 215–244.

Dunk, Jeffrey R.; Zielinski, William J.; Preisler, Haiganoush K. 2004. Predicting the occurrence of rare mollusks in northern California forests. Ecological Applications. 14(3): 713–719.

Farnsworth, George L.; Pollock, Kenneth H.; Nichols, James D.; et al. 2002. A removal model for estimating detection probabilities from point-count surveys. The Auk. 119(2): 414–425.

Lancia, Richard. A.; Nichols, James D.; Pollock, Kenneth H. 1994. Estimating the number of animals in wildlife populations. In: Research and management techniques for wildlife habitats. Bethesda, MD: The Wildlife Society: 215–253.

MacKenzie, Darryl I.; Nichols, James D.; Lachman, Gideon B.; *et al.* 2002. Estimating site occupancy when detection probabilities are less than one. Ecology. 83(8): 2248–2255.

MacKenzie, Darryl I.; Nichols, James D.; Hines, James E.; *et al.* 2003. Estimating site occupancy, colonization and local extinction probabilities when a species is not detected with certainty. Ecology. 84(8): 2200–2207.

Manley, Patricia N.; Zielinski, William J.; Schlesinger, Matthew D.; Mori, Sylvia R. 2004. Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. Ecological Applications. 14(1): 296–310.

Nichols, James D.; Conroy, Michael J. 1996. Techniques for estimating abundance and species richness. In: Measuring and monitoring biological diversity: standard methods for mammals. Washington, DC: Smithsonian Institution Press: 177–234.

Nichols, James D.; Boulinier, Thierry; Hines, James E.; *et al.* 1998a. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology. 12(6): 1390–1398.

Nichols, James D.; Boulinier, Thierry; Hines, James E.; *et al.* 1998b. Estimating rates of local species extinction, colonization and turnover in animal communities. Ecological Applications. 8(4): 1213–1225.

Nichols, James D.; Hines, James E.; Sauer, John R.; *et al.* 2000. A double-observer approach for estimating detection probability and abundance from point counts. The Auk. 117(2): 393–408.

Olsen, Anthony R.; Sedransk, Joseph; Edwards, Don; *et al.* 1999. Statistical issues for monitoring ecological and natural resources in the United States. Environmental Monitoring and Assessment. 54(1): 1–45.

Pollock, Kenneth H. 2000. Capture-recapture models. Journal of the American Statistical Association. 86: 225–238.

Pollock, Kenneth H.; Nichols, James D.; Brownie, Cavell; Hines, James E. 1990. Statistical inference for capture-recapture experiments. Wildlife Monographs. 107: 1–97.

Pollock, Kenneth H.; Nichols, James D.; Simons, Theodore R.; *et al.* 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics. 13(2): 105–119.

Powell, Larkin A.; Conroy, Michael J.; Hines, James E.; *et al.* 2000. Simultaneous use of mark-recapture and radiotelemetry to estimate survival, movement, and capture rates. Journal of Wildlife Management. 64(1): 302–313.

Royle, J. Andrew; Nichols, James D. 2003. Estimating abundance from repeated presence/absence data or point counts. Ecology. 84(3): 777–790.

Seber, George A.F. 1982. The estimation of animal abundance and related parameters. New York, NY: Macmillan. 654 p.

Thomas, Len; Buckland, Stephen T.; Burnham, Kenneth P.; *et al.* 2002. Distance sampling. In: Encyclopedia of environmetrics. Chichester, England: John Wiley & Sons: 544–552.

Thompson, William L. 2002. Towards reliable surveys: accounting for individuals present but not detected. The Auk. 119(1): 18–25.

Williams, Byron K.; Nichols, James D.; Conroy, Michael J. 2002. Analysis and management of animal populations. San Diego, CA: Academic Press. 817 p.

Yoccoz, Nigel G.; Nichols, James D.; Boulinier, Thierry. 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution. 16(8): 446–453.

The Application of FIA-based Data to Wildlife Habitat Modeling: A Comparative Study

Thomas C. Edwards, Jr.[1]; Gretchen G. Moisen[2]; Tracey S. Frescino[2]; and Randall J. Schultz[3]

Abstract.—We evaluated the capability of two types of models, one based on spatially explicit variables derived from FIA data and one using so-called traditional habitat evaluation methods, for predicting the presence of cavity-nesting bird habitat in Fishlake National Forest, Utah. Both models performed equally well, in measures of predictive accuracy, with the FIA-based model having estimates of model sensitivity. The primary advantage of using the FIA data is the ability to convert the modeled relationships to spatially explicit depictions of bird habitat.

The conservation and management of animal populations depend, in part, on accurate and parsimonious habitat models capable of identifying key components of an organism's habitat. Organisms are assumed to select a habitat that will maximize survival and reproductive success, and determining these habitat associations is essential to the understanding of the factors underlying species distribution and the maintenance of biodiversity.

The inability of small, single-scale studies to adequately explain and predict species presence, the recognition that patterns and processes are often fundamentally scale-dependent, and the desire to minimize the need for intense field sampling have all resulted in the introduction of landscape-level and hierarchical investigations into habitat selection (Lawler 1999, Lawler and Edwards 2002, Mitchell *et al.* 2001, Morris 1987, Turner *et al.* 1989, Wiens 1989, among others). Habitat selection by an individual in a population is influenced by the composition and configuration of the surrounding landscape matrix (Wiens and Milne 1989), and the incorporation of the "landscape-level" concept and technological advances allowed for habitat selection to be examined at multiple spatial scales and varying hierarchical levels (Bergin 1992, Gutzwiller and Anderson 1987, Lawler 1999, Mitchell *et al.* 2001, Saab 1999, Wiens *et al.* 1987).

Ecologists have suggested that to maximize predictive capability, habitat models need to incorporate a range of scales (Knick and Rotenberry 1995). In a management context, however, landscape-level habitat modeling is a desirable alternative to microhabitat sampling since microhabitat field sampling is often not spatially explicit, and it can be time consuming and labor intensive (Mitchell *et al.* 2001). Landscape-level modeling also allows for the study and management of the environment across large areas and in remote areas. This is a desirable goal for broad-scale wildlife management; however, these models must be applied with caution. Landscape-level habitat models must predict species presence beyond a desired accuracy, or if maximum predictive capability is the goal, landscape models must predict species distribution similarly or better than a microhabitat model or combined landscape/microhabitat model to *alone* suffice for wildlife habitat modeling.

Here we evaluate the efficacy of FIA-based data and derived information in wildlife habitat modeling. Specifically, we use FIA-derived, spatially explicit maps of several variables assumed related to the presence of wildlife. The obvious advantage of the FIA-based variables is their ability to be used in spatial extrapolation. These predictor variables are compared against more traditionally collected habitat variables (after James and Shugart 1970; hereafter "traditional model") having, perhaps, better ecological linkage to species ecology but lacking in the capability for spatial extrapolation. We test the simple hypothesis that FIA-based habitat models perform equally as well as traditional models in predictive capability. Our test species is a guild of cavity-nesting birds.

[1] Research Ecologist, U.S. Geological Survey, Utah Cooperative Research Unit, Utah State University, Logan, UT 84322–5290.
[2] Research Forester, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401.
[3] Graduate Research Assistant, Ecology Center and Department of Forest, Range and Wildlife Sciences, Utah State University, Logan, UT 84322–5230.

Methods

Our study area was the Fishlake National Forest in southern Utah at the southern extent of the Wasatch Mountains. The study area encompasses sections of four ranger districts (Richfield, Loa, Fillmore, and Beaver) spread over three general mountain ranges. The Richfield Ranger District is located on Monroe Mountain and the Eastern Ranges, the Loa Ranger District is located on the southern portion of the Eastern Ranges, and the Fillmore and Beaver Ranger District are both located on the Pahvant and Tushar Ranges, respectively (hereafter the Western Ranges). This region of Utah is characterized by high mountains (~2,000 to 4,000 m) consisting of broad, rolling plateaus, large alpine meadows, and considerable amounts of aspen (*Populus tremuloides*) forest. The winters are long and cold, and the summers are warm with frequent afternoon mountain storms and summer monsoons.

Our study species included all members of the cavity nesting bird community found to nest in aspen forests of Fishlake National Forest, Utah. The species included six primary cavity nesting birds: red-naped sapsuckers (*Sphyrapicus nuchalis*), northern flickers (red-shafted) (*Colaptes auratus*), hairy woodpeckers (*Picoides villosus*), downy woodpeckers (*Picoides pubescens*), three-toed woodpeckers (*Picoides tridactylus*), red-breasted nuthatches (*Sitta canadensis*), and six secondary cavity nesting birds: tree swallows (*Tachycineta bicolor*), violet-green swallows (*Tachycineta thalassina*), mountain chickadees (*Poecile gambeli*), mountain bluebirds (*Sialia currucoides*), western bluebirds (*Sialia mexicana*), and house wrens (*Troglodytes aedon*).

We systematically surveyed the study region for active nests of cavity nesting birds from late May until early July. A nest was considered active if it exhibited evidence of incubation, presence of eggs, presence of young, and/or feeding behavior. Due to the lack of inference available from the nest building stage, we did not include evidence of nest building as a sign of activity. If a nest was in the building stage, however, we returned to the site later to determine whether or not the nest became active. To mark the active nests, we flagged a tree that was at least 10 m away from the nest tree and recorded the distance and azimuth to the nest tree from the flagged tree. In addition, we recorded the UTM coordinates at each nest site using a global positioning system (GPS).

We measured vegetation variables within 0.04 ha surrounding active nest trees. This scale ensured complementarity of our data with the wealth of existing studies on avian habitat selection (James and Shugart 1970). Furthermore, this area is effective at characterizing the nest site since it is smaller than the average territory size for most small forest passerines (Noon 1981, Sedgwick and Knopf 1992). We measured a series of habitat measures including canopy cover, snag density, tree density, and shrub cover. These variables constituted our traditional models. FIA-derived variables were obtained from maps developed using emerging techniques (e.g., Frescino *et al.* 2001) that convert FIA data to spatially explicit representations. Variables modeled included canopy height (m), number of snags, number of live trees, and average tree height (m). Values were obtained from the FIA maps by intersecting the UTM coordinates of the nest sites with the digital FIA data. The result was two sets of observations for modeling purposes: one based on data collected within a 0.04-ha area surrounding active nest trees and the other extracted from the FIA-based maps.

We used stepwise logistic regression to model the probability of presence of nest sites based on traditional and FIA-based predictors. Although the number of nests varied depending on guild type and model type, we used all of the non-nest locations from 2001 in each habitat model. Models were evaluated using these criteria: (1) estimates of model fit based on model R^2 and the Somer's D statistic; (2) model predictive capability based on the 2001 training data; (3) predictive capability based on internal cross-validation; and (4) predictive capability based on the 2002 external model validation data. Measure of predictive capability included percent correctly classified (PCC), sensitivity, specificity, and the area under curve (AUC). The first three measures are considered threshold-dependent measures, and their values are dependent on a user-specified threshold. In our case we considered a threshold value of >0.5 to be indicative of nest site presence. AUC is threshold independent and is a measure of model predictive capability across the range of thresholds t, where $0<t<1.0$.

Results

A total of 227 nests were found during the 2001 and 2002 field seasons. Of these nests, 165 were found in 2001 and used for model building. The remaining 62 nests found in 2002 were used for model validation.

Model fit was relatively poor for both the traditional and FIA-based models (table 1), but differences between the two model forms were negligible. Percent correctly classified and AUC values were similar for both model forms for both training and cross-classified data (tables 2, 3). However, model forms differed in their sensitivity and specificity, with the FIA-based model having greater sensitivity but lower specificity. When tested with independent field data, the FIA-based model form had a somewhat lower PCC and AUC value (table 4). The same pattern in sensitivity and specificity found in the training and cross-validated data occurred in the independent data as well.

Table 1.—*Estimates of model fit for traditional and FIA-based habitat models of nest sites of cavity-nesting birds, Fishlake National Forest, Utah*

Model	R^2	Somer's D
Traditional	0.039	0.020
FIA-based	0.174	0.048

Table 2.—*Measures of model accuracy of training data for traditional and FIA-based habitat models of nest sites of cavity-nesting birds, Fishlake National Forest, Utah*

	FIA-based	Traditional
PCC	0.633	0.637
Sensitivity	.910	.593
Specificity	.239	.739
AUC	.600	.742

Table 3.—*Measures of model accuracy of cross-validated training data for traditional and FIA-based habitat models of nest sites of cavity-nesting birds, Fishlake National Forest, Utah*

	FIA-based	Traditional
PCC	0.633	0.637
Sensitivity	.910	.593
Specificity	.240	.740
AUC	.565	.720

Discussion

Forest wildlife management often requires not only understanding of the ecological reasons behind a species presence on landscapes, but also a depiction of the spatial distribution of the species. Variables suited for explaining *why* a species is found at specific locations are not necessarily the best for predicting *where* a species is located. Moreover, the types and kinds of variables associated with species presence (e.g., presence of fungal conks as an indicator of suitable trees for cavity nesting birds) are often difficult to model and map. Consequently biologists must often choose, based on management objectives, whether explaining the *why* of species presence location is more important than the *where* of species presence. Ideal models would simultaneously address both questions, but variables well suited for mapping are not the same as those suited for explanation. The results presented here indicate that habitat models for cavity nesting birds based on variables having less ecological explanatory value do as equally well in prediction as those with high ecological explanatory value. The added value to the FIA-based variables is the ability for spatial extrapolation. These spatially explicit maps can provide managers with much needed information on the spatial distribution of critical habitats.

However, use of maps of forest type and structure in wildlife management are only as accurate as the models that created the structural maps. There are several means of modeling or mapping forest structure across space, the first of which is statistical modeling. In the Uinta Mountains of Utah,

Table 4.—*Measures of model accuracy of independent data for traditional and FIA-based habitat models of nest sites of cavity-nesting birds, Fishlake National Forest, Utah*

	FIA-based	Traditional
PCC	0.557	0.703
Sensitivity	.790	.593
Specificity	.283	.830
AUC	.541	.755

Frescino *et al.* (2001) built and validated statistical models of forest presence, forest type, basal area, shrub cover, and snag density using remotely sensed imagery and a suite of environmental predictor variables (environmental gradients, temperature, precipitation, elevation, aspect, slope, and geology). The models for forest presence and forest type were 88 percent and 80 percent accurate, and an average of 62 percent of the predictions for basal area, shrub cover, and snag density fell within ~ 15 percent deviation of field values (Frescino *et al.* 2001). Such levels of accuracy are well within the margins of error for wildlife management.

The ability to predict where a species occurs and where it does not occur is vital to management decisions. Biologists must evaluate their habitat models using rigorous model validation to test the spatio-temporal accuracy of their predictions. Our results indicate, at least for the system studied here, that equally reliable models could be built using so-called traditional methods as well as new methods capable of translating FIA data into spatial representations. The advantage of latter is the clear ability to use these maps for spatial extrapolation for use in wildlife management.

Literature Cited

Bergin, T.M. 1992. Habitat selection by the western kingbird in western Nebraska: a hierarchical analysis. Condor. 94: 903–911.

Frescino, T.S.; Edwards, T.C. Jr.,; Moisen, G.G. 2001. Modelling spatially explicit forest structural variables using generalized additive models. Journal of Vegetation Science. 12: 15–26.

Gutzwiller, K.J.; Anderson, S.H. 1987. Multiscale associations between cavity-nesting birds and features of Wyoming streamside woodlands. Condor. 89: 534–548.

James, F.C.; Shugart, H.H. 1970. A quantitative method of habitat description. Audubon Field Notes. 24: 727–736.

Knick, S.T.; Rotenberry, J.T. 1995. Landscape characteristics of fragmented shrubsteppe habitats and breeding passerine birds. Conservation Biology. 9: 1059–1071.

Lawler, J.J. 1999. Modeling habitat attributes of cavity-nesting birds in the Uinta Mountains, Utah: a hierarchical approach. Logan, Utah: Utah State University. Ph.D. dissertation.

Lawler, J.J.; Edwards, T.C. Jr., 2002. Landscape patterns as predictors of nesting habitat: a test using four species of cavity-nesting birds. Landscape Ecology. 17: 233–245.

Mitchell, M.S.; Lancia, R.A.; Gerwin, J.A. 2001. Using landscape-level data to predict the distribution of birds on a managed forest: effects of scale. Ecological Applications. 11: 1692–1708.

Morris, D.W. 1987. Ecological scale and habitat use. Ecology. 68: 362–369.

Noon, B.R. 1981. Techniques for sampling avian habitats, p. 42–52. In: Capen, D.E.; ed. The use of multivariate statistics in studies of wildlife habitat. Gen. Tech. Rep. RM-87. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station: 42–52.

Saab, V. 1999. Importance of spatial scale to habitat use by breeding birds in riparian forests: a hierarchical analysis. Ecological Applications. 9: 135–151.

Sedgwick, J.A.; Knopf, F.L. 1992. Describing willow flycatcher habitats: scale perspectives and gender differences. Condor. 94: 720–733.

Turner, M.G.; O'Neill, R.V.; Gardner, G.H.; Milne, B.T. 1989. Effects of changing spatial scale on the analysis of landscape pattern. Landscape Ecology. 3: 153–162.

Wiens, J.A. 1989. Spatial scaling in ecology. Functional Ecology. 3: 385–397.

Wiens, J.A.; Milne, B.T. 1989. Scaling of 'landscapes' in landscape ecology, or, landscape ecology from a beetle's perspective. Landscape Ecology. 3: 87–96.

Wiens, J.A.; Rotenberry, J.T.; Van Horne, B. 1987. Habitat occupancy patterns of North American shrubsteppe birds: the effects of spatial scale. Oikos. 48: 132–147.

Nonnative Plants in the Inventory of Western Oregon Forests

Andrew N. Gray[1]

Abstract.—Vegetation data from the 1997 inventory of non-Federal forests in western Oregon were examined to assess the abundance of invasive nonnative plants detected by the inventory. Inventoried plants were the more common, identifiable species; composites and graminoids were underrepresented. Nonnative species were found on 1,040,000 ha (35 percent) of the non-Federal forest land in western Oregon. Abundance of most nonnatives was greater on lands that had been recently clearcut or thinned and was more closely related to stand density than climate. Most species were shade-intolerant, but shade-tolerant ivy and holly have the potential to greatly increase in abundance.

Invasion of plants and animals into new regions around the world dramatically impacts the economics of human communities, and seriously threaten native species and ecosystems (Mooney and Hobbs 2000, Vitousek *et al.* 1996). Biological invasions may be second only to land use change as causing species extinction (D'Antonio and Vitousek 1992). Competition from nonnative plants is a ubiquitous problem in most agricultural areas, and can also jeopardize the success of forest plantations and the availability of forage on rangeland (e.g., Crompton *et al.* 1988, Randall and Rejmánek 1993).

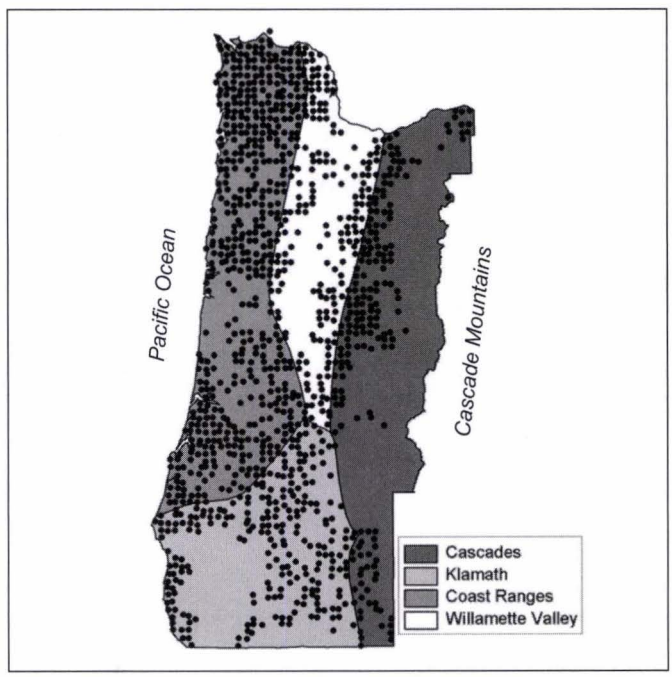
Public agencies in the Pacific Northwest are struggling to evaluate and control the spread of invasive, nonnative plant species. The Oregon Department of Agriculture maintains lists of the most threatening exotic plants and has guidelines for controlling them (Oregon Department of Agriculture 2001). However, there have been few assessments of the distribution and rate of spread of many of these species. Most published information about the distribution of individual species is derived from collected locations of specimens in herbaria, or from general descriptions of their association with plant communities published in floras. The objective of the study reported here was to assess the abundance of nonnative plant species recorded during an inventory of non-Federal forests in western Oregon and to examine the effects of environment and management on their abundance.

Methods

Inventory Design

Forest land in western Oregon was inventoried from 1995 to 1997 by the USDA Forest Service's Forest Inventory and Analysis (FIA) unit in the Pacific Northwest Research Station. The study area consisted of the counties of western Oregon, generally extending west from the crest of the Cascade Mountains to the Pacific Ocean, and south from the Columbia

Figure 1.–*Map of western Oregon study area, showing boundaries of ecoregions (Bailey 1980) and approximate locations of forestland FIA plots.*



[1]Research Ecologist, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331. Phone: 541–750–7252; e-mail: agray01@fs.fed.us.

River to the California border. The area has a broad range of climate and vegetation, ranging from the *Picea sitchensis* Zone along the moist, warm coastline, to the *Tsuga heterophylla* Zone in the cooler Coast Ranges and the lower Cascades, to *Quercus garryana* communities in the dry, warm interior valleys (Franklin and Dyrness 1973). This range of conditions can be stratified into four main ecoregions: Coast Range, Cascades, Willamette Valley, and Klamath (Bailey 1980) (fig. 1).

The field plot sample design consisted of a systematic grid on a 5.4-km spacing (1 point per 3,000 ha). Inventory data from land managed by the Bureau of Land Management and the USDA Forest Service were excluded from this study because they were not collected by FIA and did not have comparable information on vegetation composition. Field plots were installed by FIA at all grid locations on remaining lands that met the criteria for "forest land" (i.e., land area ≥0.4 ha in size that had, or had previously had, 10 percent canopy cover by trees and was not primarily managed for nonforest land use). A total of 1,127 plots were sampled.

At each forest sample grid location, five subplots were installed (or remeasured) over a 2-ha area at 66-m intervals. Understory vegetation was sampled on 5-m radius plots around each subplot center. All species of shrubs, and of trees <2.5-cm diameter at breast height (d.b.h.) ("seedlings") were recorded; "herb" species (forbs, graminoids, and ferns) were recorded if present at ≥3 percent cover. Species' cover was estimated in 5-percent increments. Trees were identified by species and sampled more intensively than other plants. Trees ≤12.5-cm d.b.h. were measured on 2.35-m fixed radius plots at each subplot point, and larger diameter trees were sampled with a variable-radius, 7-m basal area factor prism, up to a maximum distance of 17 m. Diameter, height, species, crown class, and other attributes were recorded for all sample trees. Additional measurements of tree canopy cover were made along three 17-m transects on each subplot. Trees were assigned to as many as three canopy layers, and the portions of the transect covered (using vertical projection from crown edges) by each species were recorded. Other data collected on FIA plots (e.g., root disease, snags, and down wood) were not used in this study.

The systematically placed subplots often sampled different forest types, stand age classes, or land types, termed "condition classes" (e.g., clearcut vs. pole stand, pasture vs. forest). Obvious differences in vegetation types were distinguished and mapped in the field, and all collected data were identified to the condition class in which they were found.

Analysis

There were 545 plant species recorded at the 1,127 sample grid locations that were forested. To reliably examine the distribution of a given species, it was first necessary to select only those species that had a high probability of being detected by most field crews. For each species, an estimate was made of the ability of most of the crew members to identify it if encountered (low, medium, or high) based on personal experience with plant identification in the region, and corroboration with an experienced field crew member (Erica Hanson, personal communication). Some species were combined at the genus level based on similar ecological attributes and/or difficulty of identification. Of the

Table 1.—*Plant species not native to western Oregon selected for analysis, and their frequency in the FIA inventory*

Species	Code	Common name	Number of plots	Percent frequency
Cirsium spp.	CIRSI	thistle	102	9.1
Cytisus scoparius	CYSC4	Scotch broom	55	4.9
Digitalis purpurea	DIPU	foxglove	57	5.1
Hedera helix	HEHE	English ivy	9	0.8
Hypericum perforatum	HYPE	St. John's wort	32	2.8
Ilex aquifolium	ILAQ80	English holly	41	3.6
Rubus discolor	RUDI2	Himalayan blackberry	264	23.4
Rubus laciniatus	RULA	cutleaf blackberry	148	13.1

species recorded, all 37 tree species were judged to have a high to medium probability of being identified by most crews; 84 shrubs, 50 forbs, and 0 graminoids were judged to meet the same criteria (there was no specific training for identification of graminoids for this inventory). These species were then matched with a national list of nonnative species (USDA 2002), resulting in a final list of eight species examined in this study (table 1).

Summaries of species frequency and abundance were calculated for the study area and by ecoregion. Additional analyses of species frequency in relation to environment and management history were done at the stand level. Independent variables were various potentially important site, stand, and climatic attributes. Site attributes consisted of elevation, aspect (linearized to reflect solar heating with the coversion cosine(aspect-45)), percent slope, potential annual direct radiation (from equations in Oke (1987)), and topographic moisture index (Parker 1982). Stand attributes included percent tree canopy cover, proportion of cover in hardwoods, basal area per hectare of all trees and of shade-tolerants only, quadratic mean diameter, stand size class, and stand age. Climatic variables included mean annual and mean summer precipitation, mean annual and mean summer temperature, mean annual range in temperature, and a moisture stress index of summer temperature divided by summer precipitation, all derived from the PRISM climatic interpolation model (Daly *et al.* 1994).

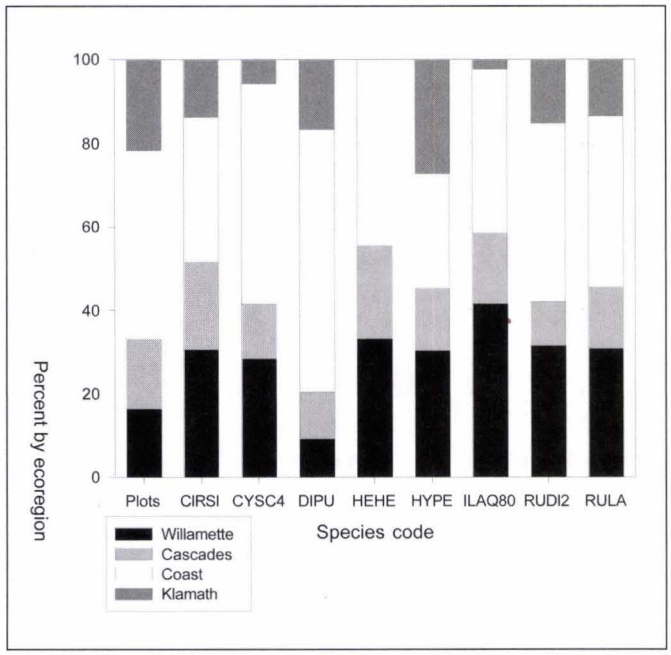
Logistic regression was used to examine species-distribution patterns where the dependent variable was the odds ratio, expressed as the number of subplots with a species over the number of subplots sampled within a stand on a plot (GENMOD procedure, SAS Institute (1999)). The regression analyses could not accommodate multiple conditions ("stands") on the same plot because many of the independent, plot variables would be identical for different stands. Therefore, only the largest condition on each plot was included in the analyses. A modified stepwise procedure was used by running the logistic regressions for each species on all independent variables individually, and then building the model with the strongest variable and assessing the strength of additional (uncorrelated) variables. Pearson correlation coefficients among the independent variables were used to help guide model development. Residuals were examined and plotted against the next variable in the model to avoid inclusion of spurious relationships.

Additional analyses were used to evaluate the effect of logging on the presence of the selected species. Two types of logging were evaluated: stand-replacing harvest (clearcutting, usually with no retention of trees from previous stand), and partial harvest (primarily commercial thinning). The former was assessed by comparing seedling/ sapling stands with a mean tree d.b.h. ≤12.5 cm with stands with larger trees. Thinning was assessed for pole- and sawtimber-sized stands (d.b.h. >12.5 cm) by comparing stands that had been partially harvested in the 10 years before measurement to stands that had not. To test the effect of logging, the best logistic regression model for a species was developed using the approach described above but without including stand density or stand size attributes, and then applying a least-square means test on either stand size or thinning (Ramsey and Schafer 1997).

Results

Overall, 450 of the 1,127 stands (40 percent), and 1,726 of the 4,924 vegetation subplots (35 percent) had one or more of the selected weed species. Applying the latter percentage to the

Figure 2.–*Proportion of sample plots by ecoregion, compared to the proportion of sample plots where each nonnative species was found, by ecoregion. Species codes are defined in table 1.*



area of non-Federal forestland in western Oregon (Azuma *et al.* 2002) suggests these species were present on 1,040,000 ha of non-Federal forestland in western Oregon. Himalayan blackberry (*Rubus discolor* Weihe & Nees) was the most abundant of the nonnative species, having been found on almost one-quarter of the plots and across most of the study area (table 1). On the other hand, English ivy (*Hedera helix* L.) was the least abundant of the nonnative species and was primarily found close to Portland and other urban areas.

The proportion of sample plots by ecoregion was usually different from the proportion of sample plots where nonnative species were found (fig. 2). All of the nonnative species, except for foxglove (*Digitalis purpurea* L.), were more common in the Willamette Valley ecoregion than in the other ecoregions. Foxglove was particularly common in the Coast Range ecoregion. The Klamath ecoregion had fewer plots with Scotch broom (*Cytisus scoparius* (L.) Link), English ivy, or English holly (*Ilex aquifolium* L.).

Stand density was strongly associated with the frequencies of all of the nonnative species, but the most significant variables, and their signs, differed among species (table 2). Increasing tree canopy diminished the frequency of thistles (*Cirsium* P. Mill.), Scotch broom, St. John's wort (*Hypericum perforatum* L.), and cutleaf blackberry (*Rubus laciniatus* Willd.). In addition, the greater the basal area the less the frequency of foxglove and Himalayan blackberry. Increasing stand density increased the frequency of English ivy and English holly, with the most important variables being canopy cover and quadratic mean diameter, respectively.

Climatic and site variables were also strongly associated with the frequency of most species. Increasing annual precipitation reduced the frequency of thistle, but increased the frequency of foxglove. There was a positive relation between annual range in temperature and frequency of St. John's wort. For the rest of the species, frequency diminished with increasing elevation. Although elevation is a complex climatic gradient, across this population of plots (where elevation ranged from 0 to 1433 m) the correlations between elevation and climatic variables were not high: r=-0.50 for annual temperature, and r=-0.11 for annual precipitation.

All the species were significantly affected by harvest activity, except for English ivy (possibly due to low sample size for this species). Most species were at least twice as frequent (close to 10 times more for thistle) in seedling-sapling stands (d.b.h. ≤12.5 cm) than in pole and sawtimber stands (d.b.h. >12.5 cm) (fig. 3). The reverse was true for English holly, and English ivy was only found in the larger size classes. Thinning the larger size classes within the 10 years before inventory also significantly affected frequency of most of the species. Except for English ivy, species were two to four times more likely to be found in thinned stands than in unthinned stands of the same size class.

Discussion

This analysis of recent forest inventory data in western Oregon indicates that nonnative species have become an important part of the flora in these forests, having been detected on 40 percent

Table 2.—*Variables selected for final logistic regression models of frequency for each nonnative species on FIA plots, the direction of the effect (as sign, positive or negative), and the strength of the variable (only F-statistics are shown; P values for all variables were <0.0005)*

Species code	Variable	Sign	$F_{(1,1124)}$
CIRSI	Tree cover	–	613.13
	Annual precipitation		31.28
CYSC4	Elevation	–	76.85
	Tree cover	–	70.4
	Aspect (cosine)	–	30.89
DIPU	Basal area	–	293.19
	Annual precipitation	+	172.57
	Topographic moisture	–	11.53
HEHE	Elevation	–	257.13
	Tree cover	+	102.82
HYPE	Tree cover	–	131.34
	Annual temperature range	+	50.49
ILAQ80	Elevation	–	124.45
	Quadratic mean diameter	+	51.52
RUDI	Elevation	–	242.33
	Basal area	–	139.34
RULA	Elevation	–	128.22
	Tree cover	–	77.29

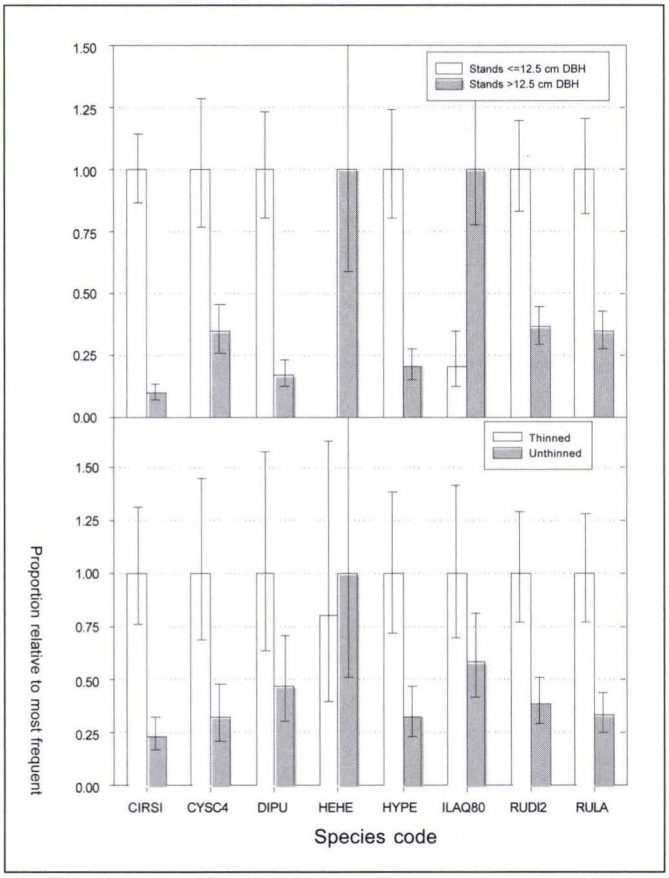
of the plots on non-Federal forest lands. It is unknown, however, if a complete botanical inventory on these plots (i.e., geared to sample species that were difficult to identify or that were regionally uncommon) would have revealed a greater prevalence of nonnative species. Many composites and grasses are hard to identify, even if encountered during peak phenology. For example, spotted knapweed (*Centaurea maculosa* auct. non Lam.) and quakgrass (*Elymus repens* (L.) Gould) are nonnative species on the Oregon Department of Agriculture (ODA) noxious weed list (Oregon Department of Agriculture 2001) but have not been detected yet on the FIA plots. It is hoped that the planned implementation of full species sampling on a subset (one of every 16) of the standard FIA plots (known as "forest health monitoring plots") in Oregon beginning in 2004 will provide information on the abundance of many more species than is currently available.

Most of the nonnative plants in this study were more likely to be found in the Willamette Valley ecoregion than in the other ecoregions of western Oregon. Many of the species were more affected by elevation than by any climatic variable. These results suggest that the presence of these plants on forest land is strongly affected by proximity to urban and agricultural settings, where anecdotal evidence indicates they are often locally abundant; unfortunately there is no comparable inventory of vegetation on nonforest land to assess this quantitatively.

Shade tolerance was an important trait determining species distributions. Six of the eight species in this study declined with increasing stand density, were significantly more common in seedling/sapling stands than in larger stands, and were more common in thinned stands than in unthinned stands. While thistle, Scotch broom, St. John's wort, and cutleaf blackberry appeared to be more sensitive to canopy cover than other measures of stand density, foxglove and Himalayan blackberry appeared to be more sensitive to basal area. Because basal area increases slower than canopy cover, foxglove and Himalayan blackberry may be more shade-tolerant than the other four species. Because they are most common during early stages of succession, these six species may be ephemeral on forest land. Nevertheless, their prevalence in recently thinned stands indicates the presence of persistent seed banks in the soil or abundant seed sources.

Unlike the relatively shade-intolerant species, the frequency of English ivy and English holly was associated with increasing stand density, and both were more common in larger stand size classes than in seedling/sapling stands. These evergreen shrubs are common in urban areas where they are planted as ornamentals, they are shade-tolerant, and their seed is readily consumed and dispersed by native birds. The ivy is currently on the ODA weed list while the holly is not. Their ability to persist in closed-canopy stands and their lack of association with any climatic variables in this study suggest a high potential for future spread in the forests of western Oregon.

Figure 3.–*Magnitude of logging effect on the frequency of nonnative species. Estimated frequencies for each species were determined by least-square means from logistic regression models, and means and 95-percent confidence intervals were expressed as proportions of the largest value to display all species on the same scale. For each species, the upper panel compares estimated frequency between stands with mean d.b.h. ≤12.5 cm with those greater, and the lower panel compares thinned and unthinned stands for stands >12.5-cm d.b.h.. Differences were significant at the p=0.05 level for all species except HEHE. Species codes are in table 1.*



Acknowledgments

Numerous individuals on the field and analysis teams of the PNW-FIA program made this study possible. Vicente Monleon helped design the statistical analysis, and Erica Hanson helped interpret the field data. Janet Ohmann, Vic Rudis, Beth Schulz, and Jeremy Fried provided helpful reviews of an earlier version of the manuscript.

Literature Cited

Azuma, David L.; Bednar, Larry F.; Hiserote, Bruce A.; Veneklase, Charles A. 2002. Timber resource statistics for western Oregon, 1997. Resour. Bull. PNW-RB-237. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Bailey, Robert G. 1980. Description of the Ecoregions of the United States. Misc. Publ. 1391. Washington, DC: U.S. Department of Agriculture.

Crompton, C.W.; Hall, I.V.; Jensen, K.I.N.; Hildebrand, P.D. 1988. The biology of Canadian weeds, 83: *Hypericum perforatum* L. Canadian Journal of Plant Science. 68: 149–162.

D'Antonio, Carla M.; Vitousek, Peter M. 1992. Biological invasions by exotic grasses, the grass/fire cycle, and global change. Annual Review of Ecology and Systematics. 23: 63–87.

Daly, C.; Neilson, R.P.; Phillips, D.L. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology. 33: 140–158.

Franklin, J.F.; Dyrness, C.T. 1973. Natural vegetation of Oregon and Washington. Gen. Tech. Rep. PNW-8. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Mooney, H.A.; Hobbs, R.J.H. 2000. Invasive species in a changing world. Washington, DC: Island Press.

Oke, T.R. 1987. Boundary layer climates. London: Routledge.

Oregon Department of Agriculture. 2001. Oregon noxious weed strategic plan: comprehensive guide for the protection of Oregon's resources. Salem, OR: Oregon Department of Agriculture.

Parker, A.J. 1982. The topographic relative moisture index: an approach to soil-moisture assessment in mountain terrain. Physical Geography. 3: 160–168.

Ramsey, Fred L.; Schafer, Daniel W. 1997. The statistical sleuth. Belmont, CA: Wadsworth Publishing Co.

Randall, J.M.; Rejmánek, M. 1993. Interference of bull thistle (*Cirsium vulgare*) with growth of ponderosa pine (*Pinus ponderosa*) seedlings in a forest plantation. Canadian Journal of Forest Research. 23: 1507–1513.

SAS Institute Inc. 1999. SAS/STAT user's guide, version 8. Cary, NC: SAS Institute, Inc.

U.S. Department of Agriculture. 2002. The PLANTS database, Version 3.5 (http://plants.usda.gov). Baton Rouge, LA: U.S. Department of Agriculture, National Plant Data Center.

Vitousek, P.M.; D'Antonio, C.M.; Loope, L.L.; Westbrooks, R. 1996. Biological invasions as global environmental change. American Scientist. 84: 468–478.

Providing Confidence in Regional Maps in Predicting Where Nonnative Species are Invading the Forested Landscape

Dennis M. Jacobs and Victor A. Rudis[1]

Abstract.—Nonnative invasive plant species introduced to the South during the past century threaten to forest resources. Knowing their extent is important for strategic management and planning. We used U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) field observations at ground-sampled locations to model the geographic occurrence probability of forest land in the sampled region, and selected nonnative invasive species in ground-sampled forest land locations. We chose kriging to interpolate and map features geostatistically, and to portray quantitative confidence in the estimates across the sampled region.

Nonnative invasive species are invading forest land. Inventory field crews, along with State partners, now record the presence and dominant cover of selected species in annual inventories in the South (USDA 2001). Using traditional numerical procedures, analysts can determine the area of timberland that contains these species, but maps provide additional information. One is able to quickly visualize the range and distribution of invasion hot spots, initiate hypotheses for species distributions, and provide tabular information to potential stakeholders. Maps help regional administrators, managers, planners and policymakers to quickly locate where nonnative invasive species occur, but more detail is needed to permit a discussion of strategies for managing nonnative invasive species and controlling their spread.

Simple dot maps of the approximate sample locations show the distribution of forest land as well as individual species, e.g., Beltz and Bertelson (1990). That same location information also may be used to interpolate the occurrence of these species between plot locations by generating a probability surface with a finer resolution than the original spacing. We report on interpolation procedures used in Rudis and Jacobs (in preparation) to map Japanese honeysuckle (*Lonicera japonica* Thunb.), privet (*Ligustrum* spp.), multiflora rose (*Rosa multiflora* Thunb. ex Murray), tree-of-heaven (*Ailanthus altissima* [Mill.] Swingle), kudzu (*Pueraria. montana* [Lour] Merr. var. *lobata* [Willd.] Maesen & S. Almeida), melaleuca (*Melaleuca quinquenerva* [Cav.] S. T. Blake), and royal paulownia (*Paulownia tomentosa* [Thunb.] Sieb. Zucc. ex Steud.) in selected States of the South.

Methods

We used plot coordinates that were nominally correct to within 800 m of the actual plot location. Land use at field plots and selected nonnative invasive species' presence and absence values on ground-sampled forest locations were data variables. We chose the survey years 1988 to 1995 as data were complete and species selection procedures were consistent within the then three FIA survey regions. For calculating forest land probability, State surveys included South Central States—Alabama 1990, Arkansas 1995, Louisiana 1991, Mississippi 1994, east Oklahoma 1993, Tennessee 1989, east Texas 1992; Southeastern States—Florida 1995, Georgia 1989, North Carolina 1990, South Carolina 1993, Virginia 1992; and Kentucky 1988.

We entered the selected data variables into a geostatistical software program (GS+ (Robertson 2000)). Calculations involved two essential steps. The first step was to assess the optimum range (the distance at which pairs of plot samples no longer influence one another and are statistically independent) and best-fit parameters from the best model (the model that minimizes the residual sum of squares while also retaining a high R-square value from the available data). The second step was to apply the parameters of the best-fit model to generate a grid surface through a routine called a kriging system (Krige 1951, Matheron 1963). We selected 1 km as the standard grid cell size and limited the range to the 16 plots closest to the 1-km grid cell.

Kriging variance provides data to produce a companion map showing reliability of the data for each grid cell value in the kriged map. These kriging variance values are mapped in

[1] Research Foresters, Southern Research Station, Forest Inventory and Analysis Unit, Starkville, MS 39760–0928. E-mail: djacobs@fs.fed.us and vrudis@fs.fed.us.

conjunction with the cell values generated during the kriging routine. A higher number of plots within a close radial distance and the shorter the distance from the grid cell to each of the surrounding plots provide lower kriging variance values to show the better quality of these data. Areas of higher kriging variance may be compared to the same areas within the forest probability map to determine which areas are suspect to a high degree of error. These geographic areas of high error are then blocked out of the map to prevent the portrayal of potentially erroneous information.

Results

The kriging variance grid (fig. 1) showed the difference in density of the plot locations across the South through the portrayal of the kriging variance value for each grid cell. Geographic areas containing high numbers of plots in close proximity to one another exhibited low kriging variance for grid cells near those plots. Where plots were less dense or unevenly spaced, the kriging variance was slightly higher for neighboring grid cells.

For Southeastern States (Florida, Georgia, North Carolina, South Carolina, and Virginia), the Southeastern Coastal Plain contained roughly one plot per 1,100 ha; the Piedmont one plot per 1,400 ha; and the Mountains, one plot per 1,900 ha. Clumping of plots, as seen in the Mountains, produced a spotty effect, i.e., a patchwork of high and low kriging variance. Kentucky was an excellent example. The western portion averaged fewer plots than any of the other States in the South. However, in eastern Kentucky, the national forest land averaged a higher proportion of plots.

The South Central States (Alabama, Arkansas, Louisiana, Mississippi, eastern Oklahoma, Tennessee, and eastern Texas) had nearly uniform plot spacing due to the historical layout of plots on a 4.8-km x 4.8-km sample grid, i.e., one plot per 2,300 ha. Also, there were no field plots in western Oklahoma and western Texas and along predominantly swampland portions of Coastal Plain counties in Louisiana and Texas. This western portion of the map shows a uniform kriging variance across the South Central region with one easily seen exception: Plot spacing within the region's small national forests was denser, thus providing for lower kriging variance. Another exception, light areas within the map not easily seen, was along county lines. In such cases, the kriging variance map illustrates a somewhat

Figure 1.—*Kriging variance, Southern United States, 1988-1995. (South Central—Alabama 1990, Arkansas 1995, Louisiana 1991, Mississippi 1994, east Oklahoma 1993, Tennessee 1989, east Texas 1992; Southeastern—Florida 1995, Georgia 1989, North Carolina 1990, South Carolina 1993, and Virginia 1992; and Kentucky 1988.)*

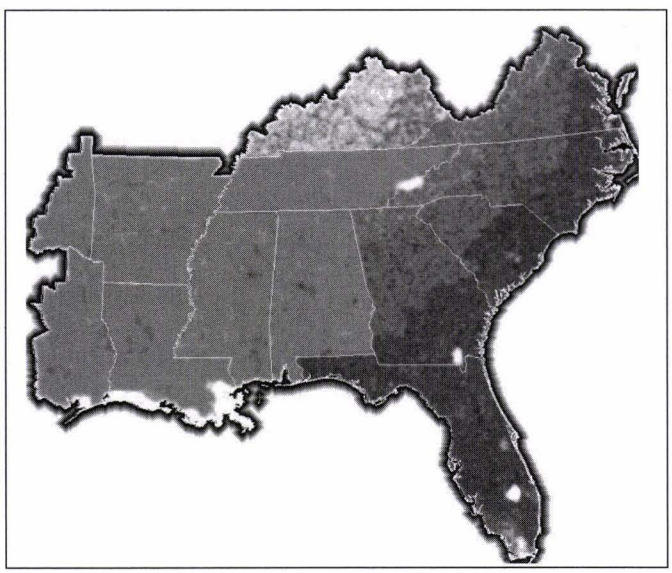
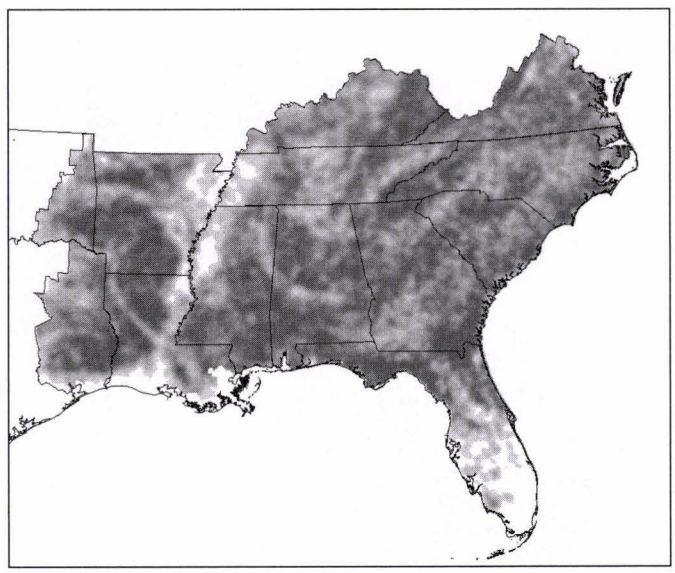


Figure 2. —*Forest land area from ground-sampled locations, Southern United States, 1988-1995. (South Central—Alabama 1990, Arkansas 1995, Louisiana 1991, Mississippi 1994, east Oklahoma 1993, Tennessee 1989, east Texas 1992; Southeastern—Florida 1995, Georgia 1989, North Carolina 1990, South Carolina 1993, and Virginia 1992; and Kentucky 1988.)*



larger spacing than the suggested 4.8-km distance for laying out the original plot grid.

Interpolation outside the grid of plots resulted in extremely high kriging variance values. Likewise, interior holes of no data are visible where ground-sampled information in areas reserved from timber production was unavailable for this study (e.g., Everglades, Great Smoky Mountains, and Okefenokee Swamp).

The forest probability map (fig. 2) portrays densely and sparsely forested areas of the South with a gradient from dark gray to light gray. Sparsely forested areas include south Florida, the Mississippi Alluvial Valley, the Arkansas River and Red River, Blackbelt Prairies, and Interstate Highway corridors. This map provides a method to block out the species grid surface layer to the grid cells having a minimum threshold forest land probability.

Discussion and Conclusion

Future plans for the location of plots include dropping some plot locations and adding others to form a regular grid network to provide for a more uniform statistical kriging variance across the entire South. As FIA and State partners implement the new wave of forest inventory plot designs across all States, more data will be available for modeling nonnative invasive species. When completed, occurrence maps will become available for detecting both the current status and change in distribution of nonnative invasive species.

New FIA procedures include refining all spatial data with more precise location information obtained from GPS units, which will improve the utility of the final map for linkage with other georeferenced data sources, and various other uses. At the same time, spatial accuracy also could be improved, especially in predominantly nonforest regions, with sampling of selected attributes on plots categorized as agricultural or urban land, but that contain forest land elements such as trees and other attributes of interest, i.e., nonnative invasive species.

Literature Cited

Beltz, Roy C.; Bertelson, Daniel F. 1990. Distribution maps for Midsouth tree species. Resour. Bull. SO-151. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 56 p.

Krige, Daniel G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of Chem. Metall and Min. Soc. South Africa. 52(6): 119–139.

Matheron, Georges. 1963. Principles of geostatistics. Economic Geology. 58: 1246–1266.

Robertson, G.P. 2000. GS+: geostatistics for the environmental sciences, version 5. Plainwell, MI: Gamma Design Software. 200 p. + CDROM.

Rudis, Victor A.; Jacobs, Dennis M. (In preparation). Regional status and change in selected non-native invasive plants of southeastern United States forests: results from 1984-1997 surveys.

U.S. Department of Agriculture. 2001 (updated August 30, 2002). Forest inventory and analysis, Southern Research Station field guide, Volume 1: field data collection for phase 2 plots, version 1.62. http://www.srs.fs.fed.us/fia/manual/p2manual.htm.

Measuring Tree Seedlings and Associated Understory Vegetation in Pennsylvania's Forests

William H. McWilliams[1], Todd W. Bowersox[2], Patrick H. Brose[3], Daniel A. Devlin[4], James C. Finley[2], Kurt W. Gottschalk[5], Steve Horsley[5], Susan L. King[1], Brian M. LaPoint[1], Tonya W. Lister[1], Larry H. McCormick[2], Gary W. Miller[5], Charles T. Scott[1], Harry Steele[3], Kim C. Steiner[2], Susan L. Stout[3], James A. Westfall[1], and Robert L. White[6]

Abstract.—The Northeastern Research Station's Forest Inventory and Analysis (NE-FIA) unit is conducting the Pennsylvania Regeneration Study (PRS) to evaluate composition and abundance of tree seedlings and associated vegetation. Sampling methods for the PRS were tested and developed in a pilot study to determine the appropriate number of 2-m microplots needed to capture variability in seedling abundance. The findings resulted in a decision to use one 2-m fixed-radius microplot per 7.3-m fixed-radius subplot of the NE-FIA design. Preliminary results indicate that one-half to two-thirds of the region's forests would require remedial treatment if preferred species are the management objective.

Forest inventory data are being used to monitor understory communities as part of the inventory of Pennsylvania by the Northeastern Research Station's Forest Inventory and Analysis (NE-FIA) unit (McWilliams *et al.* 2002). The primary objective of the landscape-level Pennsylvania Regeneration Study (PRS) is to determine the composition and abundance of tree seedlings and associated understory vegetation. The PRS is part of a larger research initiative by cooperating institutions to develop site- and species-specific stocking guidelines and other management criteria for the range of forest systems in the State. The results of a pilot study to test and evaluate sampling methods for tree seedlings and understory communities are presented along with preliminary results from the first year of data collection.

Methods

Study Region

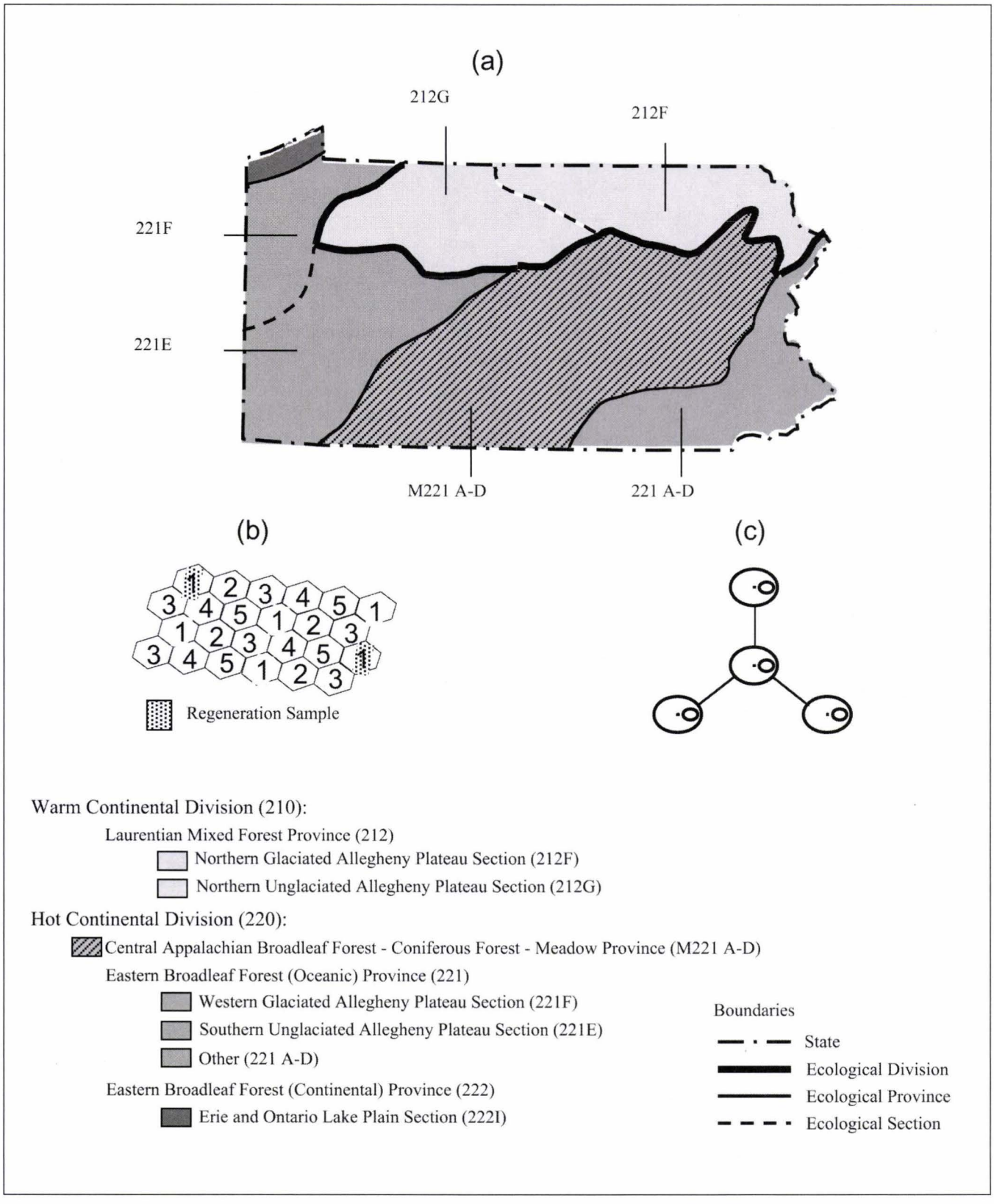
The PRS region consists of the entire State and is excellent for regeneration measurements and assessments (fig. 1a), as complicated forest associations abound: mixed mesophytic in the southwest (Braun 1985); mixed oak throughout but concentrated in the Central Appalachians; Allegheny and northern hardwoods along the northern tier; coniferous systems mixed throughout; and several other cover types (see Fike 1999). Actual species composition and structure vary greatly due to interrelated factors such as topographic location, land use and disturbance history, anthropogenic forces, and geographic differences. An overpopulation of white-tailed deer (*Odocoileus virginianus* Zimmerman) that has devastated regeneration over vast areas adds a particularly complex factor to this mix (McWilliams *et al.* 1995). Current deer populations are well above the thresholds for healthy understory development (deCalesta and Stout 1997).

Determining the Number of Microplots to Measure

Sampling methods for the PRS were tested and developed from a pilot study using a subset of NE-FIA sample locations during the 2000 field season (McWilliams *et al.* 2001). Sample plots occupy 2,400-ha hexagons that mosaic the State. Because the NE-FIA sample is measured over 5 years, 20 percent of the sample locations are measured each year in an "interpenetrating" fashion; that is, no plots are measured in two adjacent hexagons in a given year. Regeneration was measured during the leaf-on season; the interpenetrating concept (fig. 1b) was

[1] U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA, 19073. Phone: 610–557–4050; fax: 610–557–4095; e-mail: wmcwilliams@fs.fed.us.
[2] The Pennsylvania State University, School of Forest Resources, University Park, PA.
[3] U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Irvine, PA.
[4] The Pennsylvania Department of Conservation and Natural Resources, Bureau of Forestry, Harrisburg, PA.
[5] U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Morgantown, WV.
[6] Allegheny National Forest, Warren, PA.

Figure 1.—*Ecoregions of the study region (a), systematic interpenetrating sample design (b), and sample location layout (c), Pennsylvania.*



used. Each year, about 300 regeneration plots are measured. The NE-FIA sample location consists of four 7.3-m fixed-radius subplots spaced 36.5 m apart with a 2-m fixed-radius microplot used for saplings (fig. 1c).

One objective of the pilot study was to determine the appropriate number of 2-m microplots needed to determine in situ variability in seedling abundance. A model was fit for every number and spatial combination of 16 microplots—the maximum that would fit onto the sample location, or four microplots per subplot.

A relative-variance function relates the plot and sampling design variables to the attribute of interest. A relative-variance function for a single plot of varying size (z) takes the form (Smith 1938):

$$R_i(z) = b_{0i} z^{b_{1i}} \quad (1)$$

where:

b_{1i} = a negative exponent relating the area sampled to the relative-variance of an attribute i such as species

b_{0i} = a coefficient

As b_{1i} approaches 0, little information is gained by increasing plot size, whereas as b_{1i} approaches -1, increasing the plot size provides new information. Scott (1993) extended Smith's formula to include the case of multiple subplots:

$$R_i(m, \bar{d}, z) = b_{0i} m^{b_{1i}} \bar{d}^{b_{2i}} z^{b_{3i}} \quad (2)$$

where:

m = the number of subplots

$\bar{d}$ = averaged paired distance between subplots

z = subplot size

In this application, the distance between the subplots and the size of the plots is fixed. Therefore, the only variable is the number of plots. The relative-variance was replaced by its square root, the coefficient of variation (CV).

$$CV_i(m) = b_{0i} m^{b_{1i}} \quad (3)$$

Using nonlinear regression, the coefficients b_{0i} and b_{1i} converged to 1 and 0 respectively for any combination of species, and subplots. Linearizing the coefficient of variation equation, by taking the natural log of both sides allowed a differentiation of the coefficients.

$$\ln(CV_i(m)) = \ln(b_{0i}) + b_{1i} \ln(m) \quad (4)$$

The intercepts varied by species, but the slopes ranged from -0.6 to -0.16. Figure 2 shows the plotted function for all species combined. Individual species had similar curves. Since the curve was flat for four or more microplots, the decision is to use one 2-m microplot per 7.3-m subplot.

Sample Design

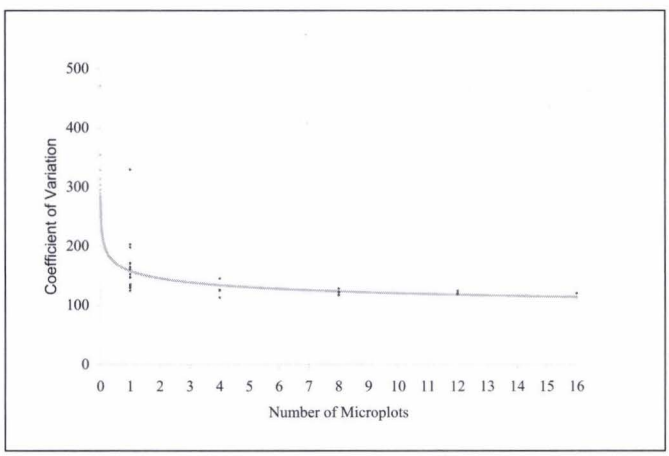
The overall nested plot design follows protocols used by Marquis (1994). A focus session with field staff following the pilot-study fieldwork resulted in suggestions for modifying tally procedures. For example, the number of tree-seedling height classes was reduced from eight to six without sacrificing scientific utility. The final design consists of a tally of all fully established seedlings (less than 2.5 cm in diameter) by species, source, and height class. Seedling source includes stump sprout, other seedling, and a "competitive" category for large-seeded deciduous species. The minimum threshold of 1.9-cm root-collar diameter for competitive status was based on Brose and Van Lear's (1998) findings for long-term stem survival. Microplot variables also include standard NE-FIA sapling measurements, presence of a large tree, and site limitations. Percentage cover of associated understory vegetation was estimated by species using the larger subplot. Marquis (1994) found that the 7.3-m size captured the variation of fern, grass, and other herbaceous vegetation. Associated understory vegetation was tallied using standard FIA codes for woody shrub species and three groups for other life forms: fern, grass, other herbaceous.

Preliminary Results

Indicators of Regenerative Capacity

Indicators used to analyze advance tree seedlings were developed to provide results that span a range of stocking that

Figure 2.—*Coefficient of variation for numbers of tree seedlings as a function of numbers of microplots used.*



reflects both standard guidelines (Gingrich 1967, Sander *et al.* 1976) and those for the high deer impact-conditions of Pennsylvania (Marquis and Bjorkbom 1982). The standard guideline for acceptable stocking is 25 seedlings per 2-m microplot versus 100 for high deer-impact conditions. Each sample tree is weighted by height class before the thresholds are applied:

Height Class	Weight
5.1 cm to 14.7 cm	1
14.7 cm to 0.3 m	1
0.3 m to 0.9 m	2
0.9 m to 1.5 m	20
1.5 m to 3.1 m	50
Greater than 3.1 m	50

Any combination of weighted stems that meets or exceeds the minimum number required is considered adequate stocking. For example, one seedling from 1.5 to 3.1 meters in height represents 50 seedlings. The indicators also used the tally of saplings (2.5 to 12.5 centimeters in diameter) to fully account for understory tree stocking. The results were partitioned by species groupings that reflect a range of management objectives: preferred, commercial, or woody (McWilliams *et al.* 1995).

Advance Tree-seedling Component

Applying the stocking thresholds to the sample data provides estimates of the proportion of forest that met or did not meet accepted silvicultural guidelines for advance tree-seedling stocking (table 1). The sample data were filtered to include only forested sample locations within the range of stocking where silvicultural guidelines indicate sufficient light for tree-seedling establishment (from 40- to 75-percent stocked with overstory trees). The findings indicate from one-third to one-half of the region's forests would need some form of remedial treatment if commercially acceptable species are the management objective; one-half to two-thirds require remedial treatment if preferred species are desired. Estimates for the indicators were lower for the Laurentian Mixed Forest Province and the Eastern Broadleaf Forest (Oceanic) Province (East) than for the Central Appalachian Broadleaf-Coniferous Forest-Meadow Province and Eastern Broadleaf Forest (Oceanic) Province (West) (Bailey 1995). Data from future samples should reveal additional spatial information, for example, test results of ecoregions and deer management zones for detectable differences in regenerative capacity; and data specific to natural and managed systems, advance- and post-disturbance regeneration, and composition of understory communities.

Associated Understory Vegetation

Tree seedlings and associated understory vegetation compete for growing space (Lorimer *et al.* 1994). Using percentage cover as a surrogate for growing space allows us to compare cover for samples that did and did not meet the tree-seedling stocking thresholds. Results for the two stocking guidelines by vegetative component are shown in table 2. Samples that did not meet the thresholds had more growing space allocated to associated understory vegetation than those that did. Fern was particularly opportunistic. The most common ferns in Pennsylvania, rhizomous, are not preferred deer food, and quickly spread across the forest floor in the absence of competition for available light.

Conclusions

The PRS results are commonly cited in policy discussions within Pennsylvania's environmental community because the implications for forest management are controversial. These include significantly reducing the State's doe herd, installing and maintaining deer fencing, applying herbicides and other control measures, and introducing prescribed fire in areas where species such as *Quercus* spp. are desired future stand cohorts. The PRS sampling protocols and indicators are useful for characterizing understory vegetation. Future work will be directed toward refining existing methods with the study team focusing on reviewing and expanding specific indicators to address a wider range of questions.

The detailed understory measurements collected in this study can be used to address additional research questions. Extensions include developing models for prospective vegetational changes based on overstory-understory relationships, gaining insight into differences between advance- and post-disturbance regeneration,

additional indicator development, and improving stocking guidelines for managed and unmanaged eastern hardwood forests. Understory measurements also will improve estimates of understory biomass and carbon by vegetational component.

Literature Cited

Bailey, Robert G. 1995. Description of the ecoregions of the United States. 2d ed. Misc. Publ. 1391 (rev.). Washington, DC: U.S. Department of Agriculture, Forest Service. 108 p.

Table 1.—*Percentage of samples that met regeneration indicator thresholds by species group, stocking guideline, and ecoregion, Pennsylvania, 2001*

Species group	Stocking guideline	Ecoregion[a]				
		All	EBF-W	PLAT	C APP	EBF-E
Preferred	Standard	48	58	45	44	50
	High deer	31	45	24	30	22
Commercial	Standard	66	69	65	64	61[b]
	High deer	51	57	46	52	39
Woody	Standard	72	77	68	72	67[b]
	High deer	56	66	47	58	50

[a] EBF-W=Eastern Broadleaf Forest (Oceanic) Province - West; PLAT=Laurentian Mixed Forest Province (Plateau); C APP=Central Appalachian Broadleaf - Coniferous Forest - Meadow Province; EBF-E=Eastern Broadleaf Forest (Oceanic) Province - East.
[b] Indicates nonsignificant differences between standard and high deer guidelines for H_0: $P_s > P_{hd}$ at the 95% confidence level.

Table 2.—*Mean percentage cover by threshold status, vegetative component, stocking guideline, and species group, Pennsylvania, 2001*

Threshold status	Vegetative Component[a]							
	F	FGH	SB	All	F	FGH	SB	All
	Standard				High deer			
				Preferred				
Met	23[b]	21[b]	17[b]	18[b]	20	21[b]	16[b]	18[b]
Not met	28	25	16	22	27	24	16	21
				Commercial				
Met	22	20	16[b]	18	20	19	16[b]	18
Not met	31	29	17	24	31	27	16	23
				Woody				
Met	23	20	16[b]	18	19	19	16[b]	18
Not met	32	29	18	25	33	27	17	23

[a] F=fern; FGH=fern, grass, and other herbaceous; SB=woody shrubs and vines.
[b] Indicates nonsignificant differences between vegetative components for H_0: $P_{NOT} > P_{MET}$ at the 95% confidence level.

Braun, E. Lucy. 1985. Deciduous forests of eastern North America. New York, NY: Hafner Press. 596 p.

Brose, Patrick H.; Van Lear, David H. 1998. Responses of hardwood advance regeneration to seasonal prescribed fires in oak-dominated shelterwood stands. Canadian Journal of Forest Research. 28: 331–339.

deCalesta, David S.; Stout, Susan L. 1997. Relative deer density and sustainability: a conceptual framework for integrating deer management with ecosystem management. Wildlife Society Bulletin. 25(2): 252–258.

Fike, Jean. 1999. Terrestrial and palustrine plant communities of Pennsylvania. Publ. 8140-bk-dcnr-1128. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Conservation and Natural Resources, Bureau of Forestry. 86 p.

Gingrich, Samuel F. 1967. Measuring and evaluating stocking and stand density in upland hardwood forests in the central states. Forest Science. 13(1): 38–52.

Lorimer, Craig G.; Chapman, Jonathon W.; Lambert, William D. 1994. Tall understory vegetation as a factor in the poor development of oak seedlings beneath mature stands. Journal of Ecology. 82: 227–237.

Marquis, David A., ed. 1994. Quantitative silviculture for hardwood forests of the Alleghenies. Gen. Tech. Rep. NE-183. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 376 p.

Marquis, David A.; Bjorkbom, John C. 1982. Guidelines for evaluating regeneration before and after clearcutting Allegheny hardwoods. Res. Note NE-307. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 4 p.

McWilliams, William H.; King, Susan L.; Scott, Charles T. 2001. Assessing regeneration adequacy in Pennsylvania's forests: a pilot study. In: Reams, Gregory L.; McRoberts, Ronald, E.; Van Deusen, Paul C., eds. Proceedings, 2d annual Forest Inventory and Analysis symposium; 2000 October 17-18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 119–122.

McWilliams, William H.; Stout, Susan L.; Bowersox, Todd W.; McCormick, Larry H. 1995. Advance tree-seedling regeneration and herbaceous cover in Pennsylvania forests. Northern Journal of Applied Forestry. 12(4): 187–191.

McWilliams, William H.; Alerich, Carol A.; Devline, David A.; *et al.* 2002. Annual inventory report for Pennsylvania's forests: results from the first two years. Resour. Bull. NE-156. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 71 p.

Sander, Ivan L.; Johnson, Paul S.; Watt, Richard F. 1976. A guide for evaluating the adequacy of oak advance reproduction. Gen. Tech. Rep. NC-23. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 7 p.

Scott, Charles T. 1993. Optimal design of a plot cluster for monitoring. In: Renolls, Keith; Gertner, George, eds. The optimal design of forest experiments and forest surveys. Proceedings of IUFRO S.4.11 conference; 1991 September 10-14; London, UK. London, UK: University of Greenwich: 230–242.

Smith, H. Fairfield. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. Journal of Agricultural Science. 28(1): 1–23.

Linking Soils and Down Woody Material Inventories for Cohesive Assessments of Ecosystem Carbon Pools

Katherine O'Neill[1], Christopher Woodall[1], Michael Amacher[2], and Geoffrey Holden[1]

Abstract.—The Soils and Down Woody Materials (DWM) indicators collected by the Forest Inventory and Analysis program provide the only data available for nationally consistent monitoring of carbon storage in soils, the forest floor, and down woody debris. However, these indicators were developed and implemented separately, resulting in field methods and compilation procedures that overlap and are not entirely compatible. Here we outline an initial approach for combining carbon estimates from the soil and DWM indicators, highlight the potential limitations of this approach, and discuss future research or protocol changes that may improve regional C estimation.

Down woody materials (DWM), the forest floor, and the upper soil horizons have been identified as critical reservoirs for carbon over time scales ranging from decades to centuries. Sequestering carbon in these pools has been proposed as a way to offset CO_2 emissions from fossil fuels. As a result, monitoring changes in carbon storage is currently required by several national and international agreements including the Montreal Process Criteria and Indicators of Sustainable Forest Management. However, no data on carbon storage in DWM exist for many forests and current estimates of soil carbon storage are largely based on static soil maps that do not reflect changes resulting from differences in land-use or management practices. In addition, many soil maps are biased toward agricultural systems and may not fully account for carbon stored in the forest floor and upper mineral horizons. Measuring these carbon reservoirs and the rates at which carbon accumulates and decomposes is critical for constraining carbon budgets in forests.

Forest Inventory and Analysis (FIA) Detection Monitoring plots (Phase 3) are the only nationally consistent data available for monitoring changes in carbon storage in forest soils, the forest floor, and woody debris. However, the Soil and DWM indicators were developed and implemented separately without reference to one another, resulting in field methods and compilation procedures that overlap and, in some cases, are not compatible for carbon accounting. This paper presents an initial approach for integrating carbon storage estimates from soil and DWM inventories to provide a cohesive, dynamic assessment of forest floor and soil carbon pools. At this early stage of implementation, there were insufficient data to provide meaningful estimates of carbon storage at the regional level. Instead, the focus of this analysis is to identify the strengths and potential limitations of combining the carbon accounting approaches for these two indicators in order to streamline data collection and compilation and improve future carbon estimates.

Methods

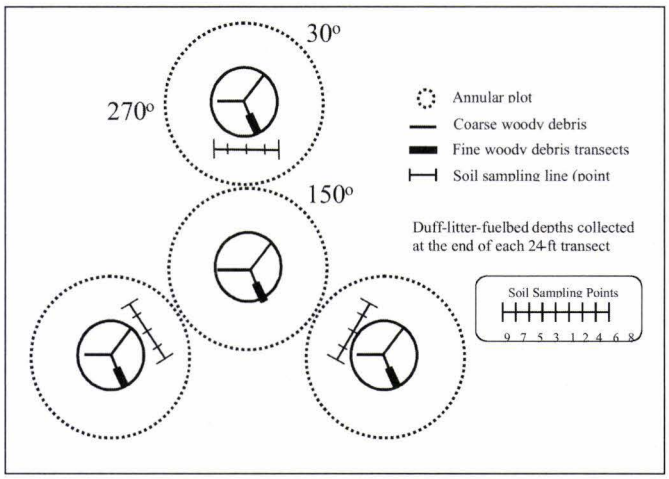
Field Methods

Soil chemical and physical properties were assessed by collecting of soil samples, which were then analyzed at a regional laboratory. Soil samples were collected within the annular plot along soil sampling lines adjacent to subplots 2, 3, and 4 (fig. 1). During the first visit to a plot, soil samples were collected at the point denoted as Soil Visit #1. On subsequent visits to a plot, soil sampling sites visit #2 or larger will be sampled. The soil sampling sites were spaced at 10-ft intervals alternating on opposite sides of soil sampling site number 1.

Samples were collected from the forest floor (subplots 2, 3, and 4) and underlying mineral soil layers (subplot 2). Forest floors were sampled after measuring the thickness at the north, south, east, and west edges of a sampling frame of known area. Once the forest floor had been removed, mineral and organic

[1] Research Soil Scientist, Research Forester, and GIS Analyst, respectively, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108.
[2] Research Soil Scientist, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Logan, UT.

Figure 1.—*Diagram of the soil and DWM sampling design.*



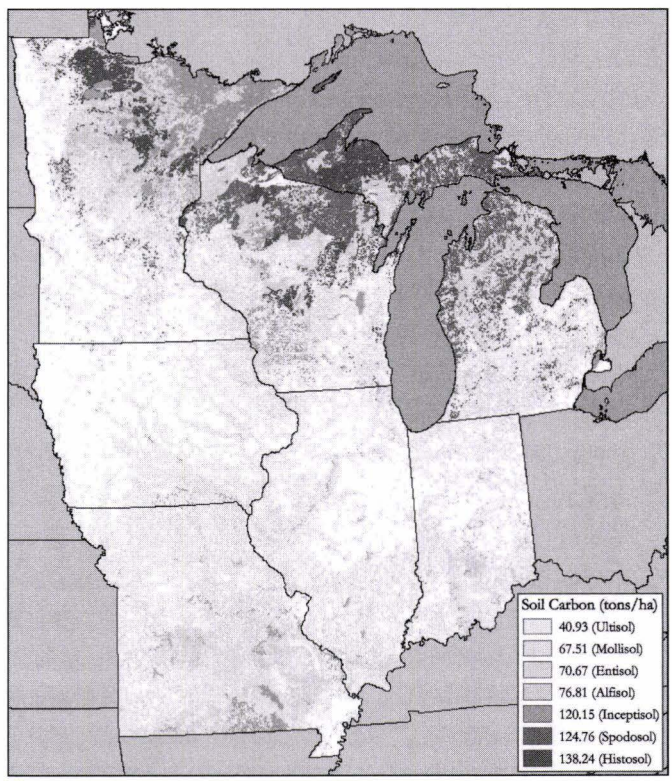
soils were sampled volumetrically by collecting bulk density cores from two depths: 0 to 10 cm and 10 to 20 cm. In organic soils, samples are collected from the litter layer and the 0 to 10 cm and 10 to 20 cm organic layers. If the soil could not be sampled at the designated sampling point due to trampling or an obstruction (e.g., boulder, tree, standing water), the sampling point was relocated to any location within a radius of 5 ft. A maximum of five samples (three forest floor and two mineral soil samples from a single core) were collected per plot. The samples were then sent to one of three regional laboratories where the carbon concentration (%) and bulk density (g cm^{-3}) were determined analytically (Palmer *et al.* 2001).

DWM were sampled using a planar-intersect method in which three 24-ft transect lines were established on each subplot (fig. 1). Coarse woody debris (larger than 3 inches in diameter) was sampled along the entire length of the transect. Crews recorded the length, diameter of the large and small ends, and the decay class of each CWD piece to allow for estimating volume. Down woody materials smaller than 3 inches in diameter, or fine woody debris (FWD), were divided into three size classes (0.0-0.25, 0.25 to 1.0, and 1.0 to 3.0 inches) and sampled along sub-sections of the transect lines. For FWD, only the number of pieces crossing the transect line was tallied. Duff, litter, and fuelbed depths were measured at a single point at the end of each transect line for a total of 12 measurements on each FIA plot.

Carbon Estimation Procedures

Carbon estimation for the soils indicator consisted of three parts (table 1). First, carbon concentration and bulk density data determined in the laboratory were combined to produce estimates of carbon storage within a single soil sample. The compilation differs slightly for mineral and forest floor samples due to the different number of samples collected for each soil layer. Next, data from each soil layer (forest floor, 0-10 cm, 10-20 cm) were aggregated within genetic soil mapping units using a digitized soil map. For this study, we used State soil survey data from the NRCS STATSGO (State Soil Geographic) database mapped according to soil order. Plot estimates of carbon storage for each layer were overlain onto a digital coverage of soil orders. The mean value of all plots that fell within a given soil order was then used to develop a spatial coverage of carbon storage within a given soil layer. Finally, the mean carbon storage estimates within each soil layer were summed to

Figure 2.—*Soil carbon storage in tons ha^{-1} (1999–2001 data). Areal carbon estimates for each soil layer were aggregated to the soil order taxonomic level using NRCS STATSGO data. The mean values for each soil order were then summed across all soil layers. Data were then masked using a forest/ nonforest map from Phase 1 of the FIA inventory. Map is not based on a full panel of data and carbon estimates are preliminary.*



provide a mean estimate of carbon storage in the upper 20 cm for a given map unit (fig. 2).

In this approach, bulk density is necessary for converting percent C concentration data to an aerial (tons ha^{-1}) basis. However, volumetric sampling for bulk density was not added to the soils protocol until 2000. In addition, not all samples collected since 2000 have a bulk density value due to the difficulty of using the impact driven corer in certain soil types. In order to use laboratory data from these plots, we applied a bulk density value estimated from STATSGO. This was done by calculating a depth- and spatially weighted average bulk density for

Figure 3.—*Carbon storage in DWM in tons ha^{-1} (2001) data. Plot level estimates of C storage were interpolated using ordinary kriging with an exponential model. Data were then masked using a forest/nonforest map from Phase 1 of the FIA inventory. Map is based on a single year of data and carbon estimates are preliminary.*

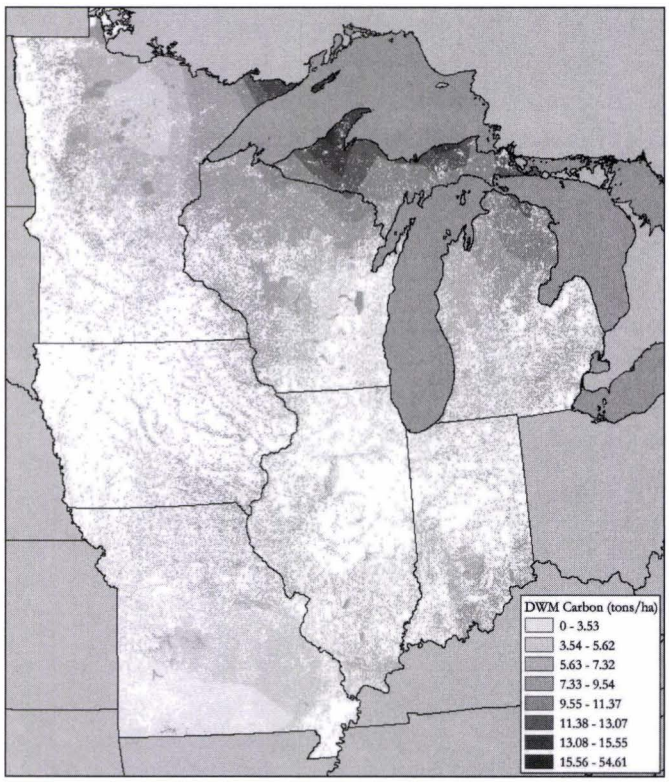


Figure 4.—*Carbon estimates in DWM, the forest floor, and the upper 20 cm of the mineral/organic soil. Data layers from figures 1 and 2 were converted to a raster image and then combined by summing across 250-m cells. Map is not based on a full panel of data and carbon estimates are preliminary.*

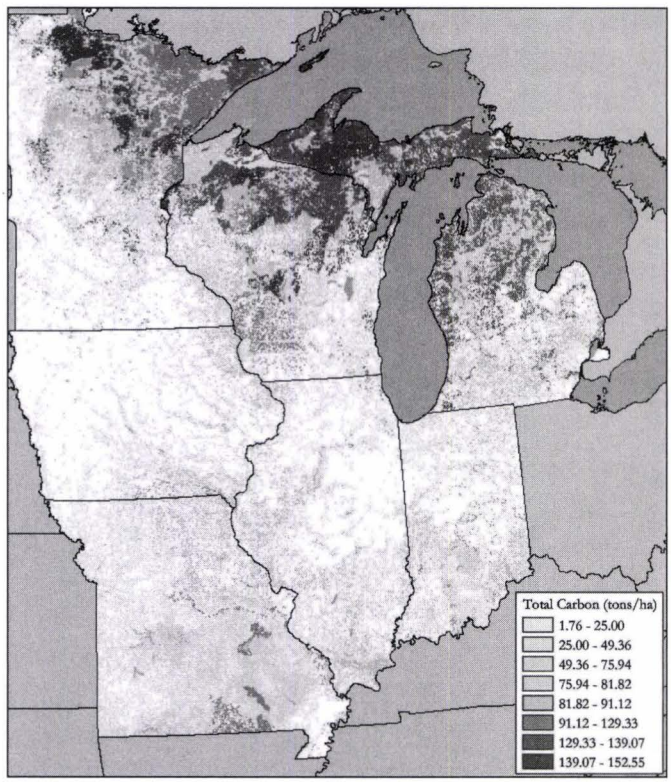


Table 1.—*Compilation and estimation procedures for the soil indicator*

	Estimate level	Units	Equation
Plot (1 sample)	Mineral Soil (0–10 cm; 10–20 cm)	g C cm^{-2}	Bulk density (g cm^{-3}) x thickness (cm) x %C
Plot (3 samples)	Mean forest floor	g C cm^{-2}	$\frac{[\text{Oven dry weight (g)/Area}(cm^{-2}) \times \%C]}{\#\text{ of samples}}$
Map Unit	Layer (forest floor, 0–10 cm, 10–20 cm)	tons C ha^{-1}	Mean of plot level estimates within a soil taxonomic unit
Map Unit	Total (forest floor –20 cm)	tons C ha^{-1}	Summation of layer C across all soil layers

each map, overlaying the FIA plots, and extracting the mean bulk density value at those locations. Carbon estimates were limited to forest soils by using a forest/non-forest map based on the Phase-1 National Land Cover Database (NLCD) imagery (Vogelmann *et al.* 2001) (fig. 2). Regional carbon estimates were based on 1,198 soil samples from 467 plots collected from 1999 to 2001.

Carbon estimates for DWM were determined at the plot level. Transect data for each size class were combined through a series of algorithms and models into plot level of estimates of volume per acre. The algorithms used for these calculations are provided in Woodall and Lutes (2005). Carbon contents were then determined by multiplying biomass estimates by a conversion factor (Birdsey *et al.* 1992, Waddell 2002). Birdsey *et al.* (1992) provide conversion factors for both softwood (0.521) and hardwood species (0.491). However, since species data were not collected on the smaller DWM size classes, we applied the mean value of these two conversion factors (0.506). Carbon storage for < 0.25 inch size class was subtracted from plot totals because this pool is also measured as part of the forest floor sample in the soils indicator. Regional estimates of DWM carbon were interpolated from plot estimates using ordinary kriging with an exponential model (fig. 3). Carbon estimates for soils and DWM were then combined by converting the spatial data layers for these indicators into a raster format and then summing across 250-m cells. The legend intervals were rescaled and then the final product was masked using a forest/nonforest map from Phase 1 of the FIA inventory (fig. 4).

Results and Discussion

Comparison of Field Methods

The key differences between the two indicators in terms of sampling design were: (1) the samples were collected on different portions of the plot; (2) the methods were designed for aggregation at different levels; (3) for the soil indicator, both the number of samples and the method used to collect these samples can vary from plot to plot; and, (4) the smallest size class of DWM (<0.25 in diameter) was also sampled as part of the forest floor sample in the soil indicator and must be subtracted from the DWM estimate to avoid double counting.

Comparison of Estimation Procedures

Carbon estimation procedures for soils and DWM differ in a number of key areas that limit both the precision and the spatial scale of combined estimates (table 2). First, both carbon concentration and bulk density were determined analytically for each soil sample in a laboratory. In contrast, the carbon content of individual pieces of DWM were either modeled (CWD) or estimated (FWD) and carbon concentrations were not empirically derived, but based on published conversion factors. Second, the DWM protocol was designed to provide a plot-level estimate of variance, with three transect lines sampled on each of the four subplots. In contrast, soil estimates were based on a single set of samples per plot (forest floor, 0-10 cm, and 10-20 cm); on some plots, the difficulty of sampling prevented collecting even a single set of measurements. This requires that C estimation for soils must be aggregated at a level above the plot scale, such as the soil order. Finally, the scaling approaches for expanding data beyond the grid framework differed, with soils data aggregated using digital soil mapping data and DWM scaled to the regional level using either ecological provinces or interpolation.

Limitations and Directions for Future Research

For the soil indicator, several limitations need to be kept in mind when estimating carbon storage. Soils are sampled destructively, which means that remeasurements will reflect a component of spatial variability in addition to change over time. Destructive sampling also has important implications for quality assurance since it is difficult for field crews to assess measurement error in bulk density. Compiling soil data is complicated by the fact that plot conditions such as difficulty in accessing soil sampling points (e.g., obstructions, water on plot), the presence of nonforested conditions, or the absence of a forest floor at the sampling point, may prevent collecting all five samples on the plot and preclude estimating C storage at the plot level. In addition, not all soils can be sampled using the bulk density corer, so crews may need to use an alternate method, such as excavation, to collect a sample for chemical analysis. Samples collected using an excavation method are not volumetric and require the use of ancillary data or interpolation

Table 2.—*Comparison of field methods and estimation procedures for the soils and DWM indicators*

	Soil	DWM
Carbon concentration	Weight %C is directly determined for sample by lab analysis (Dry combustion method).	Predicted using set conversion factors from the literature. In this study, assumed to be 0.506 for all DWM pieces.
Density	Determined directly from field samples. Oven dry weight of soil sample (g) is divided by the volume that the sample was collected from (cm^3) to provide bulk density (g cm^{-3}). In cases where sample was not collected volumetrically, determined from spatially weighted mean values from digital soil survey data (STATSGO).	For CWD (> 3.0 in diameter), estimated from the length, diameter, and decay class estimate recorded in the field. For FWD, density is determined from models.
Estimation level	Methods designed for a single estimate (without variance) at the plot level. Population estimate must be determined by averaging data within a larger unit such as soil order.	Multiple measurements collected per plot. Methods designed for determining plot level mean values.
Expansion/interpolation	Scaling to the regional level requires use of ancillary data set such as ecoregion section or soil taxonomic unit (e.g., soil orders)	Scaling done by interpolation (e.g., kriging) of plot level mean values. Plot level means can also be aggregrated using ancillary data such as ecoregion section.

based on existing data in order to scale carbon concentrations to a weight per unit area basis.

Another critical issue related to sample collection is that only one mineral core was collected per plot. This prevents assessing within-plot spatial variability and may limit our ability to detect change in carbon storage over time. Collecting additional mineral cores on subplots 3 and 4 would provide an estimate of within-plot variability that would improve carbon accounting. Although additional samples would increase the cost of laboratory analysis, modifying the field collection strategy by either lengthening the sampling cycle and reducing the number of plots sampled per year or collecting additional samples only on a subset of plots may offset these costs. Alternate sampling techniques such as composite sampling, in which multiple samples are collected in the field and then combined into a single sample for laboratory analysis, may also improve estimates in a cost-effective manner and should be investigated.

For DWM, carbon accounting limitations involve post-field sampling data processing and scaling. The DWM indicator does not directly sample carbon concentrations, but rather predicts carbon content based on the inventory of down woody components. Currently, a single conversion factor is used to predict %C. This methodology needs to be refined to better incorporate the species, decay, and inherent differences among individual DWM components. Another concern is that the DWM sample methods and data compilation routines obfuscate determination of variance estimates for different DWM components within any given sample plot. For example, different transect lengths were used to sample fine and coarse woody debris, resulting in different, but unquantified, levels of confidence in estimates. Finally, to create regional maps of DWM carbon concentrations, peer-reviewed scaling techniques need to be developed. Currently, only two methods have been proposed: ancillary data layers (e.g., ecological provinces or forest types) or data interpolation (e.g., kriging or nearest neighbor).

Summary

Although the current protocols for the soil and DWM indicators were designed and implemented separately, these data are the only source of information for monitoring changes in carbon storage in woody materials and forest soils for reporting under the Montreal Process Criteria and Indicators. For this reason, it is critical that FIA take the lead in developing statistically valid methods of carbon accounting that combine data from these two indicators. This study presented an initial approach for combining carbon estimates from the soil and DWM indicators into a regional estimate of carbon storage. As additional data become available, it will become possible to compare C estimates from this approach with values reported in the literature. Meanwhile, results from this study can be used as a basis for streamlining the field and compilation procedures for these two indicators in order to facilitate better and more efficient carbon accounting in the future. Additional research is needed to integrate these estimates with carbon models developed for other phases of the FIA program.

Literature Cited

Birdsey, R.A. 1992. Carbon storage and accumulation in United States forest ecosystems. Gen. Tech. Rep. WO–59. Washington, DC: U.S. Department of Agriculture, Forest Service. 51 p.

Palmer, C.J.; Dresbach, R.; Amacher, M.C.; O'Neill, K.P. 2001. Forest inventory and analysis soil quality monitoring: manual of soil analysis methods (2001 ed.). Available online at: http://socrates.lv-hrc.nevada.edu/fia/ia/IAWeb/Soil.htm.

Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing. 67: 6, 650–661.

Waddell K.L. 2002. Sampling coarse woody debris for multiple attributes in extensive resource surveys. Ecological Indicators. 1(3): 139–153.

Woodall, C.W.; Lutes, D.C. 2005. Sensitivity analysis of down woody material data processing routines. In: McRoberts, R.E., et al. eds. 2005. Proceedings of the fourth annual forest inventory and anlysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tec. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 199–202.

Clearcutting in the South: Issues, Status, and Trends

Jacek Siry[1] and Frederick Cubbage[2]

Abstract.—Clearcutting has been the most controversial and enduring forest management issue since its widespread adoption on public land in the 1960s. Public opinion generally opposes clearcutting, but foresters and forestry firms have adopted it widely. Despite the controversy, we have little data about the extent of clearcutting by region in the South. Forest Inventory and Analysis (FIA) data indicate 5.2 million acres are harvested annually in the South, with 39 percent being clearcut. This includes 1.67 million acres in the Southeast (85 percent clearcut), and 3.52 million acres in the South Central (17 percent clearcut). Measurement discrepancies among these regions may account for some of these differences. Including seed tree and salvage cuts, about half the timber harvests in the South are made by clearcutting. The large clearcut area, especially in the more populous Southeastern States, will continue to evoke concern about harvest practices and forest management. This issue must be addressed by careful logging and attention to public concerns, safety, and esthetic considerations in forest harvesting.

The South currently provides about 63 percent of annual timber removals in the United States (Smith *et al.* 2001) and about 18 percent of the industrial roundwood harvests in the world (FAO 2002). In addition, the South is projected to provide nearly all the increases in national timber removals over the next 50 years (Haynes, in press). Increasing removals and rising investment in timber growing will encourage more intensive management practices. Few of them are as controversial as timber harvesting, particularly clearcutting. Clearcutting removes most trees in a stand at one time, and the sight of barren forestland often evokes perceptions of widespread environmental damage, fueling opposition to its use in forest management.

The division between proponents and opponents of clearcutting is marked. Our knowledge of the extent of clearcutting and its ecological and economic impacts is modest, however. Previous studies have dealt with the environmental and economic impacts of logging practices, including clearcutting, but our basic knowledge of the extent of clearcutting in the South is almost totally lacking. Accordingly, this paper briefly reviews the current issues about clearcutting and then provides up-to-date analyses of the extent of clearcutting in the U.S. South.

The Clearcutting Issue

In the late 19th to mid-20th century, exploitative and destructive timber harvesting prompted calls for, first, Federal regulation of private forestry, and later, State forest practice laws. Currently, regulatory or non-regulatory Best Management Practices (BMPs) have been developed and implemented to protect water quality during timber harvesting, and are at least partly a response to broad concerns about clearcutting. Bliss (2000) suggests that we cannot ignore public opposition to clearcutting, no matter how compelling our scientific bases or professional beliefs. Clearcutting has been a lightning rod for public opposition to forestry practices from the cut-out-and-get-out practices of the mid-1800s to the Bitterroot and Monongahela issues in the 1960s and 1970s (Gorte 1998) to virulent opposition today. A casual search of the Internet on the subject of clearcutting is illustrative. Using *Google,* a search for the word clearcutting generated 31,200 sites; adding the word South reduced this to only 10,900 sites. A nonrandom sample of those sites indicated that most were either critiques or attacks on clearcutting, scientific articles about the subject, or professional discussions of the merits of the practice.

Critics of clearcutting state that it causes ecological degradation and soil erosion, reduces water storage capacity,

[1] Jacek Siry is an Assistant Professor at the University of Georgia Warnell School of Forest Resources, Athens, GA 30602–2152, jsiry@forestry.uga.edu.
[2] Fred Cubbage is a Professor at North Carolina State University Department of Forestry, Raleigh, NC 27695–8008, fred_cubbage@ncsu.edu.

destroys wildlife habitat, loads streams and rivers with sediment, kills fish, and results in economic ruin. The Natural Resources Defense Council states that clearcutting can jeopardize an area's ecological integrity by destroying water buffer zones and habitat for insects and bacteria, removing forest carbon sinks, eliminating fish and wildlife species via erosion, removing important underground worms and fungi, causing loss of small-scale economic opportunities, and destroying esthetic values and recreational opportunities. Other environmental groups have programs and Web sites that focus on forest practices in general, and opposition to clearcutting specifically, including the Southern Environmental Law Center (2002), The Dogwood Alliance (2002), and Heartwood (2002). A wealth of other literature exists on the potential adverse effects of clearcutting on water quality, wildlife, and scenic beauty, which is too extensive to review here.

The clearcutting issue has expanded significantly in the South since the mid-1990s. The increase of timber harvesting and wood chip mills in the Southeast has increased public discontent with forestry practices. From 1997 to 1999, the governors of Tennessee, North Carolina, and Missouri formed advisory committees to study the impact of proliferating chip mills and clearcutting. In 1999, the governor of South Carolina replied to the outcries of a coalition of 30 organizations to press for a moratorium on licensing chip mills there and to initiate a study. All those studies produced balanced reviews of forest practices and their impacts; none led to major forest policy changes to date. But widespread opposition to clearcutting and wood chip mills has not abated. More recently, the North American Coalition for Christianity and Ecology (2000) and the Progressive Presbyterians (Witherspoon Society 2001) have advocated moratoriums on clearcutting, and environmental groups throughout the South and the world continue to oppose the practice.

Scientific forestry and professional organizations have extolled the merits of clearcutting, including the Society of American Foresters (2002) and most southern State forestry associations. West Virginia University (2002) publishes a good Web-based summary on clearcutting, dispelling most myths except the obvious problem that clearcuts are (temporarily) ugly. Of course, opposition to clearcuts in Maine was so strong that the State had a ballot referendum in 1997 that unsuccessfully tried to limit clearcut sizes in the State. To combat the public protest of their recent purchase of 905,000 acres in Maine, the Plum Creek Timber Company allowed access for groups to inspect the land through guided tours. An official spokesman for Plum Creek Timber offered 100 percent public access of its land, and sponsored a media event for local newspapers in an attempt to gain public support, having been previously criticized for poor forest practices based on its reputation for clearcutting. The continued importance of clearcutting is reflected in both SFI and FSC forest certification schemes, which have clearcut size limits of 120 acres on average and 40 acres in total in the South, respectively.

Southern Clearcutting Data

From analyses performed by Siry (2002) we summarized recent FIA harvesting statistics from data sets prepared for use in SOFRA assessment to estimate the annual clearcut and partially cut areas in the 12 Southern states from Texas to Tennessee to Virginia. The latest FIA survey for each State occurred in the 1990s. In the South Central region, partial cutting, seed-tree cutting, and salvage cutting were merged into one partial cutting category that corresponds to the Southeast partial cutting category. Annual averages were obtained by dividing harvested area by the number of years between FIA surveys.

Average Annual Harvest Acreage Estimates Based on FIA Data

Table 1 summarizes the annual harvest area by type of cutting in the South by State. FIA results indicate that clearcutting occurs on about 2 million acres annually in the 12 Southern States. Upland hardwood accounts for 39 percent of clearcut land and is followed by planted pine with 22 percent (table 2). The area of clearcut planted pine is probably larger, since planted pine stands with a large hardwood component are classified as oak-pine. If so, planted pine clearcut area would be similar to upland hardwood. Clearcutting is most common on nonindustrial private land, which accounts for 57 percent of harvested area (table 3). This result is as expected because nonindustrial private owners hold most of the forestland in the region.

Table 1.—*Annual timber harvest in the South by State and type of harvest*

Region/State	Forest area	Total harvest area	Type of harvest			
			Clearcut		Partial cut	
	(acres)	(acres)	(acres)	(%)	(acres)	(%)
Southeast	**85,060,000**	**1,666,000**	**1,415,000**	**85**	**251,000**	**15**
FL	14,651,000	268,000	247,000	92	21,000	8
GA	23,796,000	543,000	446,000	82	97,000	18
NC	18,710,000	316,000	265,000	84	51,000	16
SC	12,45,000	313,000	276,000	88	37,000	12
VA	15,448,000	226,000	181,000	80	45,000	20
South Central	**103,329,000**	**3,518,000**	**600,000**	**17**	**2,918,000**	**83**
AL	21,932,000	765,000	168,000	22	597,000	78
AR	18,392,000	531,000	0	0	531,000	100
LA	13,783,000	593,000	109,000	18	485,000	82
MS	18,587,000	804,000	171,000	21	633,000	79
OK	4,895,000	94,000	7,000	7	87,000	93
TN	13,965,000	229,000	60,000	26	169,000	74
TX	11,774,000	501,000	85,000	17	416,000	83
South	**188,389,000**	**5,184,000**	**2,014,000**	**39**	**3,169,000**	**61**

Table 2.—*Annual timber harvest in the South by timber type and type of harvest*

Timber Type	Region								
	Southeast			South Central			South		
	Harvest area	Clearcut	Partial cut	Harvest area	Clearcut	Partial cut	Harvest area	Clearcut	Partial cut
	acres	%	%	acres	%	%	acres	%	%
Planted pine	396,000	98	2	383,000	12	88	779,000	56	44
Natural pine	210,000	76	24	552,000	5	95	761,000	25	75
Oak pine	300,000	84	16	701,000	13	87	1,000,000	35	65
Upland Hardwood	520,000	79	21	1,455,000	25	75	1,975,000	39	61
Bottomland Hardwood	241,000	83	17	415,000	16	84	656,000	41	59
Nonstocked	0	0	0	13,000	51	49	13,000	51	49
Total harvest area	1,666,000	85	15	3,518,000	17	83	5,184,000	39	61

Table 3.—*Annual timber harvest in the South by ownership and type of harvest*

Owner	Region								
	Southeast			South Central			South		
	Harvest area	Clearcut	Partial cut	Harvest area	Clearcut	Partial cut	Harvest area	Clearcut	Partial cut
	acres	%	%	acres	%	%	acres	%	%
Public	78,000	84	16	160,000	14	86	238,000	37	63
Forest industry	434,000	94	6	990,000	17	83	1,424,000	41	59
Miscellaneous corporate	186,000	85	15	303,000	16	84	489,000	425	58
Nonindustrial private	968,000	81	19	2,065,000	17	83	3,033,000	38	62
Total harvest area	1,666,000	85	15	3,518,000	17	83	5,184,000	39	61

The area clearcut in the South grew by nearly 10 percent over the period covered by the FIA surveys. This represents a 1.4 percent annual increase from 1986 to 1993. While clearcutting increased on public, nonindustrial private, and miscellaneous corporate land, it actually decreased on forest industry land by 5 percent. The total annual clearcut area is only about 1 percent of timberland area in the region.

FIA data indicate that partial cutting is more widespread in the South than clearcutting, occurring on about 3.2 million acres annually. Partial cutting acreage has increased by 12 percent over the period covered by the FIA surveys. The total area on which harvest cuts (clearcutting and partial cutting) are carried out is about 5.2 million acres. Clearcutting was done on about 40 percent of the harvested area. Partial cutting accounted for the remaining 60 percent of the harvested land.

Data Issues

During the SOFRA review process, the clearcutting estimates based on FIA data were called too conservative and suggestions were made that clearcutting estimates should correspond to the total area harvested, including the area that was clearcut and partially cut. Although we found no support for this proposition, we examined the FIA results in greater detail and used other sources of information to develop average annual clearcutting estimates for the South.

Examination of the FIA results revealed large differences between the Southeast and the South Central. For example, in the Southeast clearcutting accounted for 85 percent of the harvested area while in the South Central partial cutting accounted for 83 percent of the harvested area (table 1). Although some of these differences may be explained by different ownership, management objectives and approaches, and local forest conditions, these factors alone do not explain such big differences in the harvest area estimates.

Another factor that could have contributed to these discrepancies is differences in harvest definitions and their interpretation by the individual FIA units in both regions as well as our assumptions concerning the development of South-wide cutting categories. We assumed that the extent of clearcutting in the Southeast is described by the harvest variable defined as the liquidation of a merchantable-size stand of timber, leaving insufficient residual stocking for a manageable stand. In the South Central, we used the clearcut variable defined as a removal of all merchantable trees. Although these two definitions appear to be similar, there were larger differences between partial cutting definitions.

In the South Central, partial cut, seed-tree and shelterwood cut, and salvage cut variables were combined into one partial cutting category. Partial cut includes all selection cuts, high-grading, diameter-limit cutting, and any other sawtimber

cutting practice that leaves a residual stand of crop trees or potential crop trees and cull trees. It does not include poletimber thinning. Seed-tree and shelterwood cuts leave a small number of crop trees to provide seed or shade to establish a new stand. Salvage cuts remove damaged or salvable dead trees.

In the Southeast, there is only one partial cut category and that includes selective cutting and high grading—the removal of selected trees of highest value from a merchantable stand of timber, leaving sufficient stocking of residual trees for a manageable stand. This category excludes commercial thinning and other stand improvements used to enhance the growth and quality of the stand).

These definitions and assumptions indicate that while the total area on which harvest (clearcut and partial cut) took place can be reliably estimated from FIA data, there could be problems in determining the precise extent of clearcutting and partial cutting. One possible solution may be developing our own clearcutting and partial cutting definitions based on initial and residual stocking as well as volume removed. Without more detailed information about relevant FIA procedures, however, it is difficult to make additional assumptions so other information sources, such as timber sales and logging surveys, would need to be used.

Greene *et al.* (1997) provide another means to check our summary of the FIA data. They surveyed nearly 6,000 private timber sales between 1988 and 1994 in Georgia and the neighboring States of Alabama, Florida, South Carolina, and Tennessee, and recorded the type of harvest used, i.e., clearcut or partial cut. The median timber sale was 85 acres. They found that clearcutting was used on 67 percent of the sales and partial cutting on 33 percent. Furthermore, sales on forest industry land used clearcutting exclusively.

The analysis of FIA data and other sources of information indicates that annual clearcut area may be higher than that based purely on FIA data and our assumptions about combining various categories of harvest cuts. Further, seed-tree cutting and salvage cuts in the South Central could be considered clearcuts. Then, South-wide, clearcutting and partial cutting would each have a 50-percent share in harvest cuts, both being used on about 2.6 million acres annually. If Greene *et al.* (1997) estimates hold for the whole South, clearcutting would be occurring on nearly 3.5 million acres and partial cutting on the remaining 1.7 million acres. This would imply that in our estimates based on FIA data too many harvested acres were classified as clearcuts in the Southeast and too many acres were classified as partial cuts in the South Central.

While clearcutting area apparently is greater than that based on our analysis of the reported FIA data, it is not likely that all harvested land was clearcut. First, partial cutting is frequently practiced in hardwood stands, and even if these stands were high-graded, sufficient trees, albeit many of poor quality, may have been left. That may be the case if the objective was to harvest sawtimber of high-value species, leaving lower grade logs and less desirable species. Second, the growing success of Best Management Practices (BMPs) also indicates that only partial cutting is practiced in these areas. These voluntary programs require that up to 50 percent of trees will be left following harvest in Streamside Management Zones (SMZs), areas adjacent to streams and lakes. Cubbage and Woodman (1993) estimated, for example, that, in Georgia, SMZs cover about 1.5 million acres or 7 percent of the State's forestland. Growing compliance and stricter requirements indicate that partial cutting is the only harvesting practice permitted and practiced on southern forestland covered by SMZs.

Conclusions

Given the best available evidence presented here, we conclude that the annual clearcut area in the South averages about 3 million acres and can vary between 2.5 and 3.5 million acres annually. The total annual average harvest area is nearly 5.2 million acres. This area increased by 14 percent during the 7-year FIA survey cycle, or about 2 percent annually.

The average area harvested annually is likely to increase in the future to meet growing demand. As total harvest volumes increase, so will the harvested area. Results of the current RPA and SOFRA assessments indicate that the South will continue to be a major timber supplier in the United States and that harvests will increase considerably (Adams 2002, Prestemon and Abt 2002). While increasing harvests will increase harvest areas, this trend will be mitigated by the growing productivity of forest plantations.

Past experience indicates that harvest volume increases are accompanied by increases in harvest areas. But the growth in harvested areas was less than proportional. Over the period covered by FIA surveys, average annual removals of growing stock increased by 26 percent while the area harvested increased only by 12 percent (Conner and Hartsell 2002). In other words, the area of harvest cuts increased only half as fast as volume harvested.

One reason for increased timber removals from less land area is the increasing productivity of southern forests, primarily of intensively managed pine plantations. Growing more timber per acre allows meeting wood demand by harvesting less timberland area. This could be important on forest industry land, which is intensively managed. Indeed, FIA data indicate that harvesting intensity as measured by the percent of area harvested has decreased on forest industry land. Greene *et al.* (1997) also found that forest industry sales averaged 59 tons per acre versus 40 tons per acre from private sales. Technical innovations, such as wood chip mills, have allowed greater volume utilization per acre as well. These factors suggest that the area of clearcuts and partial cuts in the South will continue to increase more slowly than harvest volumes.

While intensive forestry and better utilization will foster more efficiency, the use of clearcutting must be sensitive to the context of the specific intended forestry operation. The practice will remain contentious. At the very least, it is esthetically undesirable, and at least some of the environmental concerns over its use may have merit. Most of the general public dislikes clearcuts, as evidenced by the extensive Web sites, as well as by the limits on clearcut size in the industry-initiated SFI program. The practice of clearcutting must be done in an ecologically sensitive manner, adhering to Federal, State, and local environmental guidelines, as well as forest certification standards. Strict enforcement of these guidelines is also required to protect forest areas and to ensure that forest operations will continue to have reasonable freedom in the future. Continued research into the ecological, economic, and social effects of clearcutting versus other timber harvesting methods also can help clarify tradeoffs and values. As this paper suggests, perhaps half of our timber harvests in the South are made by clearcuts, with a greater share occurring in the Southeast than the South Central. Our ability to continue practicing such even-age management in the future will depend on our skill in doing it well, with minimum adverse impacts today.

Literature Cited

Adams, D. 2002. Harvest, inventory, and stumpage prices. Journal of Forestry. 100(2): 26–31.

Bliss, J. 2000. Public perceptions of clearcutting. Journal of Forestry. 98(12): 4–9.

Conner, R.; Hartsell, A. 2002. Forest area and conditions. In: Wear, D.; Greis, J., eds. Southern forest resources assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 357–402.

Cubbage, F.; Woodman, J. 1993. Impacts of streamside management zones on timber supply in Georgia. In: Proceedings, Southern forest economics workers meeting; 1993 April 23–24; Durham, NC: 130–137.

Dogwood Alliance: http://www.dogwoodaliance.org/aboutus.asp.

FAO (Food and Agriculture Organization of the United Nations). 2002. State of the world's forests, 2001. Rome, Italy: Food and Agriculture Organization of the United Nations: http://www.fao.org/fo/sofo/SOFO2001/publ-e.stm.

Gorte, R. 1998. Clearcutting in the national forests: background and overview. Congressional Research Service Report 98-917: http://www.cnie.org/NLE/CRSreports/Forests/for-21.cfm.

Greene, D.; Harris, T.; DeForest, C.; Wang, J. 1997. Harvesting cost implications of changes in the size of timber sales in Georgia. Southern Journal of Applied Forestry. 21(4): 193–198.

Haynes, R. 2002. An analysis of the timber situation in the United States: 1952–2050. Gen. Tech. Rep. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Heartwood Forest Watch Database: http://cgi.hoosier.net/cgi-bin/cgiwrap/heartwd/showAll.cgi?national.

Natural Resources Defense Council: http://www.nrdc.org/land/forests/fcut.asp.

North American Coalition for Christianity and Ecology. Southern clergy seek moratorium on clearcutting. Earthkeeping News 9(4): May/June 2000: http://www.nacce.org/2000/chattanooga.html.

Prestemon, J.; Abt, R. 2002. The southern timber market to 2040. Journal of Forestry. 100(7): 16–23.

Siry, J. 2002. Intensive timber management practices. In: Wear, D.; Greis, J., eds. Southern forest resources assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 327–340.

Smith, B.; Vissage, J.; Darr, D.; Sheffield, R. 2001. Forest resources of the United States, 1997. Gen. Tech. Rep. NC-219. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Society of American Foresters: http://www.safnet.org/archive/clearcut.htm.

Southern Environmental Law Center. Cleaning up the South's dirtiest waters: http://www.selcga.org/act_rivreov_tmdl_st.shtml.

West Virginia University Extension Service. The clearcutting controversy – myths and facts: http://www.wvu.edu/~agexten/forestry/clrcut.htm.

Witherspoon Society. Environmental and faith groups join to urge control of forest clearcutting in the Southeast: http://www/withersponsociety.org/Global/sve_southern_forests.htm.

The Effect of Data Quality on Short-term Growth Model Projections

David Gartner[1]

Abstract.—This study was designed to determine the effect of FIA's data quality on short-term growth model projections. The data from Georgia's 1996 statewide survey were used for the Southern variant of the Forest Vegetation Simulator to predict Georgia's first annual panel. The effect of several data error sources on growth modeling prediction errors was determined, including the effect of site index measurement errors. The study suggests that for tree attributes, such as volume by species-diameter class combinations, data quality will be the largest source of prediction error. For plot attributes, site index measurement errors will be the largest source of prediction error.

With the change from a periodic statewide survey to the current rotation panel system, a method of combining the data from several panels into a single estimate is needed. The current official statistic is the moving average. However, the moving average will be biased in the presence of a linear trend. Therefore, an alternative that will reduce this bias is needed. One of the alternatives being considered by the Southern Station is imputation. Previous short-interval studies (Gartner and Reams 2002) have suggested that using growth model projections will improve the imputation results. However, growth model projection errors will be incorporated into the variance of imputation results. This stimulated my interest in growth model projection errors.

Research on the propagation of measurement errors in the input data through the growth projection process has found that site index measurement errors created some of the largest variations in the predicted values (Gertner and Dzialowy 1984, Mowrer and Frayer 1986). Since I did not have much confidence in our site index estimates, I decided to empirically estimate the amount of prediction error due to different measurement errors, including site index measurement errors.

Methods

Data

The data from Georgia's 1996 statewide survey were used to predict Georgia's first annual panel. The site indices from the first panel were used in the growth model. Only plots that were completely within one condition class were used. Plots that had been harvested during the time between measurements and plots that had no trees or saplings were not used. This left 369 plots and over 9,000 trees. Even though the surveys were about 2 years apart, the actual elapsed time between measurements ranged from 0.1 to 3.6 years.

Model

The Southern variant of the Forest Vegetation Simulator (FVS) (Donnelly *et al.* 2001) from the Forest Service's Forest Management Service Center in Fort Collins, Colorado, was used to make the predictions. To incorporate the effects of the different elapsed times between measurements, predictions were made for 1, 2, and 3 years. Then the changes predicted by the growth model were multiplied by the actual elapsed time divided by the number of years in the growth projection. For example, for the plot with 3.6 years of actual elapse time, the growth model projected changes were multiplied by 3.6 and then divided by 3.0.

Effects

The study involved: 1) using the FVS growth model with the site index estimate from the first panel, 2) removing the effects of tree damage on tree growth, 3) eliminating some apparent diameter and height data problems, and 4) rerunning the FVS growth model with a range of site indices to determine which site index minimized the residual sum of squares for individual tree volumes for each plot.

Damaged trees were taken to be outliers in terms of the growth model's behavior. That is, the growth model was designed to predict growth that is uninterrupted by exogenous damage.

[1] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, Knoxville, TN 37919. Phone: 865–862–2066; e-mail: dgartner@fs.fed.us.

Damaged trees had their observed values set to their predicted values. About 4 percent of the trees had signs of damage.

To determine whether a tree had questionable data, I created an acceptance region for diameter and height growth that ranged from a maximum growth plus a measurement error to zero growth minus the measurement error. The maximum growth rates were determined by running the North Carolina State University pine plantation growth model for loblolly pine. I used 600 trees per acre and the highest site index permitted by the software, which happened to be 99 feet base at age 25 years. Then I multiplied the maximum growth rates of the quadratic mean diameter and the dominant height by 1.5. This produced a maximum diameter growth rate of 1.135 inches per year and a maximum height growth rate of 6 feet per year. I took the diameter measurement error to be 0.5 inches and the height measurement error to be 15 feet. Trees with growth data outside this region had their observed values set to the predicted value. Less than 2 percent of the trees fell outside the acceptance region

To determine the amount of growth model prediction error due to site index measurement error, I searched potential site index values to determine the site index that minimized the sum squared error for tree volume estimates for each plot. Because of the difficulty in adapting a growth model for use in standard optimization routines, I resorted to a grid search. The site indices were varied in 1-percent increments from 55 percent to 200 percent of their panel 1 values. Not all plots reached their minimum in this range, but the residual sum of squares for these plots was less than 2 percent of the total residual sum of squares for the "optimal" site indices.

The sum squared differences between the values for the 1996 statewide survey and the first panel were calculated for diameter growth, height growth, individual tree volume growth, basal area growth, plot volume growth, and plot mortality. For the tree variables, only the surviving trees were used. These sums of squares were not corrected for the means. The sizes of the effects were measured as the percent reduction in the sum of squares.

Results

Tree Variables

The growth model did a better job at predicting diameter growth than height growth or volume growth (fig 1). The effect of site index measurement errors was only 5 percent of the total growth sum of squares for diameters, about 12 percent for height, and about 9 percent for volume.

The reduction in the residual sum of squares caused by editing out probable diameter and height data errors ranged from 10 to 18 percent. For diameters and tree volumes, the diameter and height errors contributed twice the sum of squares of the site index measurement errors. However, the height data errors contributed less to the height error sum of squares than the site index measurement errors.

Plot Variables

The growth model predicts basal area growth and plot volume growth well, around 48 percent of the sum of squares for each, but not for mortality (fig. 2). The growth model predictions

Figure 1.—*Contributions of the growth model, damaged trees, data errors, and site index to the percent growth sum of squares for diameter, height, and individual tree volume.*

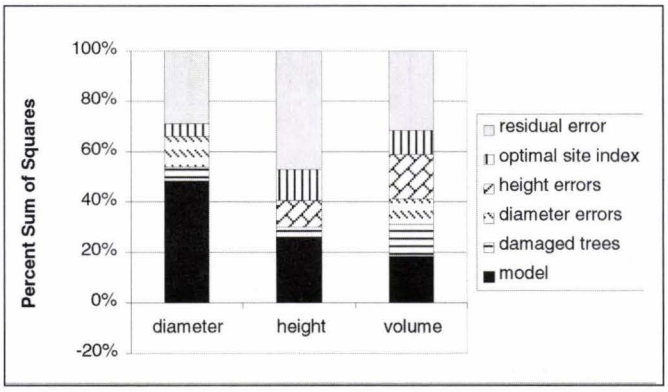
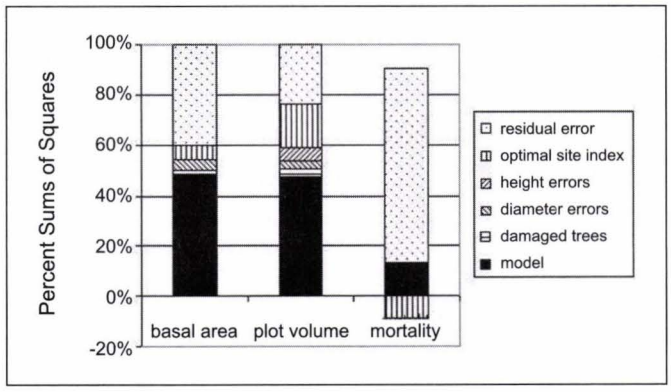


Figure 2.—*Contributions of the growth model, damaged trees, data errors, and site index to the percent growth sum of squares for basal area per acre, volume per acre, and mortality.*



using the standard site index estimates reduced the mortality sums of squares by about 16 percent. But using the optimal site index caused an 11-percent increase in the sums of squares.

The effect of site index measurement errors reaches almost 17 percent for plot volume. The effects of the diameter and height data errors are much smaller for the plot variables than for the tree variables.

Discussion

The site index that minimizes the sum squared error for individual tree volume is a function of not only the true site index, but also the growth model. Therefore, the optimal site index in this study may not be the site index as measured in the field. So the effects of the optimal site index should be considered a maximum attainable result. Decreasing the site index measurement errors by increasing the number of trees used to estimate the site index may not reduce the sum squared error shown here.

Also, the data used averaged only 2 years apart. This study may need to be repeated when data 5 years apart become available.

Dave Hyink[2] noted that other versions of the FVS model have given unusual mortality predictions with the default parameters. His experience showed that resetting the maximum-potential-basal-area parameter greatly improved the mortality predictions. This suggests that the mortality prediction function can be easily improved.

The data used for this study were measured before some of the new national data standards were implemented. One of these standards in particular requires field crews to electronically flag any observations of trees that lose more than 0.5 inches in diameter. If this new standard can prevent accidental recordings of reductions in diameter, then most sums of squares for questionable diameter data will become part of the model sum of squares. The Southern Station FIA unit's data acquisition team is implementing a similar data error check for height measurements.

Growth prediction errors are only some of the errors that will contribute to the variance of imputation results. Roesch's (1999) simulation study suggests that the greatest source of additional variation will be associated with predicting harvesting and conversions from forest to nonforest. We currently don't have any good models for predicting harvesting rates and intensities.

The long-term goal is to determine the different sources of error that contribute to the variances of imputation results and to determine the tradeoffs available to reduce these sources of variance. This study suggests that for tree attributes, such as volume by species-diameter class combinations, data quality will be the largest source of prediction error. The data acquisition band is already working on this problem. For plot attributes, site index measurement errors will be the largest source of prediction error. For combinations of plots, predicting harvesting rates and intensities will become the largest source of prediction error. This study is a small first step in determining the different sources of error that contribute the variances of imputation results, and the tradeoffs available to reduce these sources of variance.

Literature Cited

Donnelly, Dennis; Lilly, Barry; Smith, Erin. 2001. The Southern variant of the forest vegetation simulator. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Management Service Center. 61 p.

Gartner, David; Reams, Gregory. 2002. A comparison of several techniques for estimating the average volume per acre for multipanel data with missing panels. In: Reams, Gregory A.; McRoberts, Ronald E.; Van Deusen, Paul C., eds. Proceedings, 2nd Annual Forest Inventory and Analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture. Forest Service, Southern Research Station. 76–81.

Gertner, George Z.; Dzialowy, Paul J. 1984. Effects of measurement errors on an individual tree-based growth projection system. Canadian Journal of Forest Research. 14: 311–316.

Mowrer, H.T.; Frayer, W.E. 1986. Variance propagation in growth and yield projections. Canadian Journal of Forest Research. 16: 1196–1200.

Roesch, Francis. 1999. Mixed estimation for a forest survey sample design. In: Proceedings of the Section on statistics and the environment of the American Statistical Association.

[2] David Hyink, personal communication, Senior Scientific Specialist, Forest Resources Research and Engineering, Weyerhaeuser Company, Tacoma, WA 98477.

Partitioning the Uncertainty in Estimates of Mean Basal Area Obtained from 10-year Diameter Growth Model Predictions

Ronald E. McRoberts[1]

Abstract.—Uncertainty in model-based predictions of individual tree diameter growth is attributed to three sources: measurement error for predictor variables, residual variability around model predictions, and uncertainty in model parameter estimates. Monte Carlo simulations are used to propagate the uncertainty from the three sources through a set of diameter growth models to estimate the total uncertainty in 10-year predictions of mean basal area per unit area for a sample of Forest Inventory and Analysis plots. Response surface methodology is used to partition the total uncertainty by source. Of the three sources, the uncertainty in parameter estimates contributes most to the variance of the estimate of mean basal area per unit area.

The objectives of this study were threefold: (1) to obtain 10-year predictions of mean plot basal area per unit area for a sample of Forest Inventory and Analysis (FIA) plots using diameter growth models; (2) to propagate uncertainty from three sources (measurement error, residual variability around model predictions, and uncertainty in parameter estimates) through the models to estimate the total variance of mean plot basal area per unit area, and (3) to partition the total uncertainty in the mean plot basal area estimates by underlying source.

Methods

The FIA program of the USDA Forest Service has initiated an annual forest inventory system featuring measurement of a proportion of plots each year, 20 percent annually in much of the eastern United States. One approach to calculating annual inventory estimates using data obtained with the new system is to update to the current year data for plots measured in previous years and then base estimates on the updated information for all plots. If the updating procedure is sufficiently unbiased and precise, this approach is nearly as precise as using all plots but without the adverse effects of using out-of-date information. With the latter estimation approach in mind, a set of individual tree, diameter at breast height (d.b.h.) (1.37 m above ground) growth models was constructed and calibrated for use in updating FIA plot information.

The mathematical form of the d.b.h. growth models is:

$$E(\Delta DBH) = \beta_1 \exp(-\beta_2 DBH) DBH^{\beta_3} \exp\left(\sum_{j=4}^{11} \beta_j X_{j-2}\right) \qquad (1)$$

where E(.) is statistical expectation, d.b.h. is annual d.b.h. growth, the βs are parameter to be estimated, and the Xs are predictor variables in addition to d.b.h. The additional predictor variables, X_2-X_9, include a suite of tree and plot variables either measured by FIA field crews or calculated from their measurements. Tree variables include d.b.h., crown ratio (CR), and crown class (CC) at the time of the initial inventory. CR is the proportion of tree height that is in the crown, and CC is a measure of a tree's dominance in relation to adjacent trees in the same stand and is coded as follows: 1-open grown; 2-dominant; 3-codominant; 4-intermediate; and 5-overtopped (USDA FS 2001). Plot variables include latitude (LAT) and longitude (LON) of the plot center, plot basal area (BA), and physiographic class (PC). PC is a measure of site soil and water conditions that affect tree growth coded as follows: 3-xeric; 4-xeromesic; 5-mesic; 6-hydromesic; 7-hydric; and 8-bottomland (USDA FS 2001). Plot basal area in trees larger than the subject tree (BAL) is a plot variable but is calculated for each tree. BA and BAL are the sum of cross-sectional areas of live tree boles at breast height and are scaled to a per unit area basis. Details regarding calibration of model (1) are discussed by Lessard (2001).

The Annualized Inventory Database

An annualized 11-year database of plot and tree variables was constructed using the methodology described by McRoberts (2001) to provide a basis for estimating model prediction uncertainty and the total uncertainty of mean plot BA esti-

[1] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

mates. The database was constructed using measurements of forested FIA plots in Michigan, Minnesota, and Wisconsin in Bailey's eco-province 212 (Bailey *et al.* 1994) for the two most recent USDA periodic inventories in each state (Michigan 1979: Spencer and Hahn 1984; Michigan 1993: Leatherberry and Spencer 1996; Minnesota 1977: Spencer 1982; Minnesota 1990: Miles *et. al.* 1995; Wisconsin 1983: Raile 1983: Wisconsin 1996; Schmidt 1998). Because special analyses were necessary to estimate the uncertainty in the d.b.h. growth model parameters, the data were restricted to plots that included only the four most commonly occurring tree species on FIA plots in eco-province 212: red pine, jack pine, balsam fir, and quaking aspen. Thus, if any tree on a plot was any other species, the data for that plot were excluded from the database. The resulting database included information for 2,900 trees on 185 plots.

Beginning with the year 0 annual database values, the models were used to predict d.b.h. growth to obtain d.b.h. estimates each tree for years 1-10. Values of all predictor variables dependent on d.b.h.s were recalculated each year based on the d.b.h. predictions for that year. Estimates of mean plot BA and the standard error of the mean were calculated using stratified estimation (Cochran 1977) where the strata are defined by quartile categories of plot BA, and plots are assigned to strata on the basis of the year 0 plot BA. Estimates of mean plot basal area and the standard error of the mean obtained using this procedure were designated the MODEL estimates. As a standard for comparing the MODEL estimates, estimates of mean plot basal area and the standard errors of the means were calculated each year using the data in the annualized database with the same stratified estimation techniques and were designated the ANNUAL estimates.

Uncertainty in Model Predictions

Uncertainty in d.b.h. growth model predictions was attributed to three sources: uncertainty in values of predictor variables due to measurement errors, residual variability around model predictions, and uncertainty in model parameter estimates. Because of their minimal distributional and linearity requirements and because they produce reliable estimates of model prediction distributions, Monte Carlo methods were used to estimate the total uncertainty in predictions from the growth models and to propagate the uncertainties to the mean plot BA estimates. Before the simulations could be implemented, uncertainty had to be quantified for the underlying sources: measurement error for tree- and plot predictor variables, residual variability, and uncertainty in parameter estimates.

Uncertainty in Predictor Variables.—Distributions for measurement errors for the tree predictor variables were obtained from the literature. McRoberts *et al.* (1994) reported the results of a study in which 9-10 FIA field crews independently measured the same plots. They estimated a curve for describing the standard deviation of d.b.h. measurements as a function of mean d.b.h. They also reported that the distribution of ocular estimates of CR as a percentage in the 0-1 range often deviated ±0.3 around the median crew estimate. Nichols *et al.* (1991) reported that when crews returned to plots later in the same growing season to obtain second ocular estimates of CC, 80 percent of estimates were unchanged while the remaining 20 percent were evenly distributed in the two adjacent classes. Although BA and BAL are plot variables, their estimates are based on individual tree d.b.h. measurements and are also subject to d.b.h. measurement error. Uncertainty in BA and BAL was simulated by using d.b.h. measurements that incorporated simulated measurement error. Finally, because of the nonuniformity of plot soil, topographic, and vegetation conditions, PC is also subject to uncertainty due to sampling variability. However, because no empirical estimates of the sampling variability for PC were available, no uncertainty in the measurement of this variable was considered. In addition, no uncertainty was considered for the LON and LAT predictor variables.

Residual Variability.—Estimates of residual variability were obtained as by-products of calibrating the models. Residuals were assumed to follow a Gaussian distribution but with heterogeneous variances. The standard deviations of the distributions of residuals were found to be adequately described as:

$$E\left[\ln\left(\hat{\sigma}_{res}\right)\right] = \alpha_1 + \alpha_2 \ln\left(\Delta\hat{DBH}\right) \qquad (2)$$

where E(.) denotes statistical expectation, σ_{res} is the sample estimate of Φ_{res}, and $\Delta\hat{DBH}$ is predicted diameter growth from the models.

Uncertainty in Model Parameter Estimates.—Using the distributions of residual variability as previously described, distributions of model parameter estimates were obtained using a four-step Monte Carlo procedure:

1. The parameter estimates obtained from calibrating the models were used with the growth models (1) to predict d.b.h.; using predicted d.b.h., and (2), a residual was randomly selected and added to each prediction to simulate an observation of d.b.h.
2. Simulated measurement errors for d.b.h., CR, CC, and PC were obtained by randomly selecting from the appropriate distributions and adding them to the observed values to obtain simulated observations of these predictor variables; BA using the simulated d.b.h. observations was calculated for each plot, and BAL was calculated for each tree on each plot.
3. Using the simulated observations of d.b.h. from Step 1 and the simulated observations of the predictor variables from Step 2, the models were recalibrated, and the resulting parameter estimates recorded.
4. Steps 1-3 were repeated 250 times to construct a distribution of simulated parameter estimates.

Uncertainty in Model Predictions.—Estimates of mean plot BA and the standard error of the mean were obtained using a four-step Monte Carlo procedure:

1. Year 0:
    a. Each simulation was initiated by simulating measurement of all plots by adding the year 0 observed values of d.b.h., CR, and CC in the annualized database and simulated measurement errors obtained by randomly selecting values from the appropriate distributions;
    b. BA for each plot and BAL for each tree on each plot were calculated using the simulated d.b.h. observations;
    c. Mean plot BA and the standard error of the mean were calculated;
    d. A set of model parameter estimates was randomly selected from the distribution for each species.
2. Subsequent years:
    a. Current year d.b.h. for each tree was calculated as the sum of previous year's d.b.h., the model prediction of d.b.h., and a residual randomly selected from a Gaussian distribution using predicted d.b.h. and [2];
    b. BA for each plot and BAL for each tree on each plot were calculated using the simulated d.b.h. observations;
    c. Mean plot BA and the standard error of the mean were calculated and recorded;
3. Step 2 was repeated 10 times to obtain estimates of mean plot BA and the standard error of mean for years 1-10.
4. Steps 1-3 were repeated 250 times to obtain distributions of estimates of mean plot BA and the standard error of the mean for each year.

For this study, each simulation was considered a separate, independent imputation. Rubin (1987) advocates multiple completions of data sets via imputation to allow assessing the uncertainty in imputed variables and to protect against extreme results and further recommends the separate estimates be combined as follows:

$$\overline{V} = \frac{1}{m}\sum_{k=1}^{m}\overline{V}_k \qquad (3)$$

and

$$\mathrm{Var}(\overline{V}) = \frac{1}{m}\sum_{k=1}^{m}\mathrm{Var}(\overline{V}^k) + \frac{m+1}{m}\sigma^2_{\overline{V}} \qquad (4)$$

where & $\overline{V}^k$ and $\mathrm{Var}(\overline{V}^k)$ are the stratified estimates of the mean plot BA and the variance of the mean, respectively, for the k^{th} simulation, and $\sigma^2_{\overline{V}}$ is the variance among the separate estimates of mean plot BA. For this study, m=250, far greater than the m=2 or m=3 found to be adequate in unrelated studies by Rubin and Schenker (1986).

Partitioning Uncertainty

The goal in partitioning uncertainty is to quantify the contributions of uncertainties from individual sources to the uncertainty of the estimate of interest. For this study, the total variance of the model-based estimates of mean plot BA for year 10 was partitioned with respect to uncertainty from three aggregated sources: (1) measurement error, (2) residual variability around d.b.h. growth model predictions, and (3) uncertainty in parameter estimates. The uncertainties from all sources were aggregated into these three sources; i.e., measurement errors for all variables were aggregated into the single source, measurement error; residual variabilities for all species were aggregated into

the single source, residual variability; and uncertainties in all parameter estimates were aggregated into the single source, uncertainty in parameter estimates. The uncertainties for individual sources are incorporated into the simulations separately, but their contributions to the total uncertainty of the BA estimates are combined within their respective aggregated sources.

Two approaches to partitioning uncertainty are intuitive. First, the contribution to uncertainty of a single aggregated source may be estimated as the difference between the total uncertainty obtained when the uncertainties for that aggregated source are incorporated and the total uncertainty obtained when no uncertainty from any source is incorporated. This approach is denoted NONE+1. Second, the contribution of a single aggregated source may be estimated as the difference between the total uncertainty when the uncertainties for all sources are incorporated and the total uncertainty when uncertainties for all sources except the aggregated source of interest are incorporated. This approach is denoted TOTAL-1. Estimates of the contributions of individual sources obtained using the NONE+1 and the TOTAL-1 approaches are frequently biased. The bias may be seen by comparing the sums of the estimates of the contributions of all aggregated sources obtained using the NONE+1 and the TOTAL-1 approaches to the difference between the total uncertainty when uncertainties for all aggregated sources are incorporated and the total uncertainty when no uncertainty for any source is incorporated. If the estimates of the contributions from the individual sources are unbiased, the former sums should equal the latter difference. Typically they are not equal when using the NONE+1 and TOTAL-1 approaches. The bias is attributed to lack of independence among the effects of individual sources of uncertainty inherent in the simulation process.

An approach that produces independent estimates of the contributions to total uncertainty by aggregated source is based on response surface methodology (Myers 1971, Khuri and Cornell 1996). With this approach, small-order polynomials are used to describe the relationship between levels of uncertainty for underlying sources and the uncertainty of the estimate of interest. If estimates of total uncertainty are obtained for a factorial arrangement of the levels of uncertainties for the underlying sources and coded through orthogonal transformations, then a response surface may be constructed using orthogonal polynomials that produces uncorrelated coefficient estimates for first-order variables.

For each of the three sources of uncertainty, three levels of uncertainty were considered: the first level incorporated uncertainties for all individual sources corresponding to the standard deviations of the distributions of uncertainty for those sources; the second level simultaneously incorporated uncertainties for all individual sources corresponding to half the standard deviations; and the third level corresponded to no uncertainty from any component source. For the measurement error of predictor variables, the standard deviations were those obtained from the literature, and for residual variability, the standard deviations were calculated from (2). For model parameter estimates, uncertainties for the first level were incorporated in the simulations by randomly selecting from the simulated distributions of parameter estimates. For the second level, random selections were made from the simulated distributions, the deviations of these selections from the means of the distributions were calculated, and then half this deviation was added to the mean. For the third level, the means of the simulated distributions were used. Within each source, the combinations of levels of uncertainty for the individual sources are limited to three: simultaneous use of the full standard deviations for all component sources, simultaneous use of half the standard deviations for all component sources, and no uncertainty for any component source. Thus, 27 sets of simulations were conducted, one for each of the 27 combinations resulting from the three levels of uncertainty for each of the three sources.

The levels of uncertainty for each aggregated source were transformed to facilitate describing the total uncertainty of the mean plot BA estimates using orthogonal polynomials. For each aggregated source, Φ_{max} represented the first level corresponding to the full standard deviation, Φ_{min} represented the third level corresponding to no uncertainty, and Φ represented an arbitrary level. Orthogonal transformations were then applied using the coding formula of Khuri and Cornell (1996):

$$\sigma' = \frac{2\sigma - (\sigma_{min} + \sigma_{max})}{\sigma_{min} - \sigma_{max}} \qquad (5)$$

where Φ and Φ' were the untransformed and transformed codings, respectively. Although the standard deviations of the distributions of uncertainties for the individual sources differed, the transformed codings of the three levels of the uncertainties were the same for all individual sources: Φ'=1 for the first

level, $\Phi'=0$ for the second level, and $\Phi'=-1$ for the third level. Thus, the three common values, ($\Phi'=1$, $\Phi'=0$, $\Phi'=-1$) were used to describe the levels of uncertainty for an entire aggregated source. Orthogonal polynomials were based on the three values for each of three predictor variables, Φ'_1, Φ'_2, and Φ'_3, one for each source. These 27 combinations of values of 1, 0, and -1 for the three sources constituted an orthogonal design. Thus, $\mathrm{Var}(\bar{Y})$ was described using orthogonal polynomials ex-pressed using linear, quadratic, and two-way interaction terms as:

$$\mathrm{Var}(\bar{Y}) = \beta_0 + \sum_{i=1}^{3}\beta_i\sigma'_i + \sum_{i=1}^{3}\beta_{ii}\sigma'_i + \sum_{i=1}^{3}\beta_{ij}\sigma'_i\sigma'_j + \varepsilon \qquad (6)$$

where $\mathrm{Var}(\bar{Y})$ is the estimated variance of mean plot BA obtained from (4), Φ'_i is the predictor variable associated with the i^{th} source of uncertainty, β_0 is the intercept coefficient, the β_is are linear coefficients, the β_{ii}s are quadratic term coefficients, and the β_{ij}s and interaction term coefficients.

Although the estimates of the β_is are uncorrelated with each other because of the orthogonal design, the estimate of any β_i is not uncorrelated with the estimate of β_0, the estimates of the β_{ii}s, or the estimates of the β_{ij}s. Nevertheless, the coefficient estimates may be used to estimate the contribution to the total variance of the estimate of mean plot BA from the three aggregated sources and to partition the variance with respect to the contributions from those sources. The total uncertainty in the mean plot BA estimates was calculated using (6) with $\Phi'_1=\Phi'_2=\Phi'_3=1$, which corresponds to the maximum or first level of the uncertainty for all component sources. The portion of the total uncertainty attributed to the i^{th} aggregated source was estimated by setting $\Phi'_i=-1$, the minimum or third level of uncertainty for that source, and setting $\Phi'=1$, the maximum level, for the other aggregated sources, calculating the uncertainty of the mean plot BA estimate using (6), and subtracting the result from the total uncertainty estimate. This approach is analogous to the NONE+1 approach, except that it is based on predictions from (6) rather than simulated estimates. An approach analogous to the TOTAL-1 approach was also used. The estimate of uncertainty remaining after the contributions from each of the three sources have been estimated was attributed to natural variability among plots, can only be reduced by using techniques such as stratified estimation, and was designated sampling variability. Because the estimates of the contributions of aggregated sources are independent, the NONE+1 and the TOTAL-1 approaches produce identical results when used with a linear model, but do not necessarily produce identical results when the model includes quadratic and/or interaction terms.

Results

The adequacy of the 250 simulations was checked by evaluating the stability of estimates of means and standard errors of means. Plots were ordered by their variability over simulations in these coefficients of variation, and a graph of coefficients of variation versus simulation for the four plots with the greatest variability revealed that stability was achieved by approximately 100-150 simulations. Therefore, 250 simulations were deemed adequate to evaluate uncertainty.

The MODEL mean plot BA estimates tracked the ANNUAL means closely, while the MODEL standard errors were only slightly greater than the ANNUAL standard errors (table 1). The Wilcoxon Signed Ranks test (Conover 1980) detected no statistically significant differences ($\alpha = 0.05$) between the ANNUAL and the MODEL estimates of mean plot BA. The slight differences in the standard error estimates indicate that the additional uncertainty due to using the growth model predictions to predict d.b.h. introduced little additional uncertainty into the standard errors of the 10-year mean plot BA estimates.

Table 1.—*Comparisons of ANNUAL and MODEL estimates of mean plot BA*

Year	ANNUAL		MODEL	
	Mean	SE	Mean	SE
0	6.6413	0.2235	6.6413	0.2235
1	7.4136	0.2418	7.4574	0.2386
2	8.2129	0.2765	8.4553	0.2869
3	9.1482	0.3129	9.9568	0.3445
4	10.0728	0.3607	10.7690	0.4253
5	11.1123	0.4261	12.0680	0.5256
6	12.2704	0.5028	13.5109	0.6404
7	13.6350	0.5970	15.1985	0.7797
8	15.0879	0.6925	16.9822	0.9266
9	16.6195	0.7965	18.8644	1.0884
10	18.3350	0.9086	20.9654	1.2566

Table 2.—*Partitions of the variance of mean plot BA*

Aggregated source	Method			
	Observed		Response surface	
	NONE+1	TOTAL-1	LINEAR R^2=0.8902	QUADRATIC R^2=0.9999
Sampling variability (SV)	0.8902	0.8902	0.7309	0.8869
Measurement error (ME)	0.0146	0.0409	0.0227	0.0227
Residual variability (RV)	0.0043	0.0032	0.0008	0.0008
Parameter uncertainty (PU)	1.5462	1.5906	1.5682	1.5682
Subtotal 1 (ME+RV+PU)	1,5651	1.6347	1.5918	1.5917
Subtotal 2 (TOT-SV)	1.6030	1.6030	1.5918	1.5917
Total (TOT)	2.4932	2.4932	2.3227	2.4786

Of the three sources of uncertainty considered, parameter uncertainty made the greatest contribution to total uncertainty, while the contributions of measurement error and residual variability were negligible (table 2). A comparison of the Subtotal 1 and Subtotal 2 values for the NONE+1 and TOTAL-1 approaches revealed the bias inherent in the estimates of the contributions of the aggregated sources. Although the differences were not great, the LINEAR and QUADRATIC response surface models produced values that were nearly identical. Due to orthogonality, this result was expected and necessary for the LINEAR model but was an unexpected positive result for the QUADRATIC model. Based on the large R^2 =0.9999 for the QUADRATIC model, the estimates of the contributions of the aggregated sources were considered reliable.

Conclusions

Two conclusions may be drawn from this study. First, the model-based d.b.h. prediction technique had only a slight negative impact on the total uncertainty of 10-year predictions of mean plot BA. Second, among the uncertainties propagated through the model, uncertainty in the model parameter estimates made the greatest contribution to the total uncertainty in the mean plot BA estimates. Admittedly, a complete prediction system also requires techniques for predicting the survival, regeneration, and removal of trees, components that were not considered in this study.

Acknowledgments

The author gratefully acknowledges Dr. Veronica C. Lessard, National Resources Conservation Service, U.S. Department of Agriculture for assistance in developing software to implement the simulation procedures and to Dr. Christopher W. Woodall, North Central Research Station, USDA Forest Service, for assistance in implementing the growth models.

Literature Cited

Bailey, R.G.; Avers, P.E.; King, T.W.H.; McNab, W.H. 1994. Ecoregions and subregions of the United States (map). Washington, DC: U.S. Department of Agriculture, Forest Service.

Cochran, W.G. 1977. Sampling techniques. 3d ed. New York, NY: John Wiley and Sons.

Conover, W.J. 1980. Practical nonparametric statistics. 2d ed. New York, NY: John Wiley and Sons.

Khuri, A.I.; Cornell, J.A. 1996. Response surfaces: designs and analyses. New York, NY: Marcel Dekker.

Leatherberry, E.C.; Spencer, J.S., Jr. 1996. Michigan forest statistics, 1993. Resour. Bull. NC-170. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 144 p.

Lessard, V.C.; McRoberts, R.E.; Holdaway, M.R. 2001. Diameter growth models using Minnesota forest inventory and analysis data. Forest Science. 47(3): 301–310.

McRoberts, R.E. 2001. Imputation and model-based updating techniques for annual forest inventories. Forest Science. 47(3): 322–330.

McRoberts, R.E.; Hahn, J.T.; Heft, G.J.; Van Cleve, J. 1994. Variation in forest inventory estimates. Canadian Journal of Forest Research. 24: 1766–1770.

Miles, P.D.; Chen, C.M.; Leatherberry, E.C. 1995. Minnesota forest statistics, 1990, revised. Resour. Bull. NC-158. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 145 p.

Myers, R. 1971. Response surface methodology. Boston, MA: Allyn and Bacon.

Nichols, N.S.; Gregoire, T.G.; Zedaker, S.M. 1991. The reliability of tree crown classification. Canadian Journal of Forest Research. 21: 698–701.

Raile, G.K. 1983. Wisconsin forest statistics, 1983. Resour. Bull. NC-94. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 113 p.

Rubin, D.B. 1987. Multiple imputation in non-response surveys. New York, NY: John Wiley and Sons.

Rubin, D.B.; Schenker, N. 1986. Multiple imputation for interval estimation for simple random samples with ignorable nonresponse. Journal of the American Statistical Association. 81: 366–374.

Schmidt, T.L. 1998. Wisconsin forest statistics, 1996. Resour. Bull. NC-183. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 150 p.

Spencer, J.S. 1982. The fourth Minnesota forest inventory: timber volumes and projections of timber supply. Resour. Bull. NC-57. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 72 p.

Spencer, J.S.; Hahn, J.T. 1984. Michigan's fourth forest inventory: timber volumes and projections of timber supply. Resour. Bull. NC-72. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 95 p.

U.S. Department of Agriculture. 2001. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.5. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis.

Individual-tree Green Weight Equations for Loblolly Pine (*Pinus taeda* L.) Sawtimber in the Coastal Plain of Arkansas

Travis E. Posey, Paul F. Doruska, and David W. Patterson[1]

Abstract.— Loblolly pine (*Pinus taeda* L.) weight equations were developed to predict outside-bark, green bole weight to a 4-inch diameter-inside-bark (dib) top and an 8-inch dib top in southeast Arkansas. Trees were sampled from 8 different tracts over the first half of 2002: 4 tracts during winter and spring, respectively. The sampled trees ranged from 10 to 30 inches diameter at breast height (d.b.h.) and from 45 to 100 feet in total height. Parameter estimates did not differ significantly by season. The developed equations were compared with others published in the Southeast. Not surprisingly, the equations developed here outperformed the others examined for these data.

Weight scaling has become popular in the southern United States for buying and selling loblolly pine (*Pinus taeda* L.) saw logs. At most sawmills, saw logs are bought and sold exclusively by weight (primarily by the U.S. ton – 2000 pounds) to save time and money. However, sawtimber inventory volumes in south Arkansas are usually calculated in terms of Doyle board feet. It is difficult to compare the value of a stand based on a timber volume inventory ($/MBF) to the prices offered at the mill ($/U.S. ton) because of the different units involved. Volume tables developed in south Arkansas are readily available for conducting stand inventories in both cubic feet and board feet. However, there are few publicly available equations or tables that accurately report saw log weight for this region. The need exists for calculating inventory results by weight rather than volume for this region. We undertook a project to:

1. Develop loblolly pine sawtimber-sized tree weight equations using trees sampled in southeast Arkansas;
2. Determine the differences, if any, in bole weight equation parameters between winter and spring; and
3. Compare the equations developed with published equations from the Southeastern United States.

Methods

Procedure

All study sites were located in southeast Arkansas on land owned by Plum Creek Timber Company. Some 155 saw log-sized loblolly pine trees were sampled in eight stands. Eighty-one trees were weighed during February 2002, and 74 were weighed in May 2002. This allowed seasonal differences in weight equations to be examined. Table 1 summarizes the information for stands sampled during winter and spring.

Each stand was visited twice, once before and once after harvest. The first visit consisted of locating and measuring the trees that would later be weighed. In each stand, 20 loblolly pine trees (≥ 10" d.b.h. class) were selected by systematic random sampling. Several measurements including d.b.h. (inches), number of 17-foot logs, total height (feet), and bark thickness

Table 1.—*Mean and standard deviation (in parentheses) of inventory data by season*

Attribute	Winter 2002	Spring 2002
Basal area (sq. ft/acre)	76.0 (8.1)	69.5 (14.5)
Pine trees/acre	58.8 (6.2)	67.9 (18.9)
Site index (ft. base age 25)	65.6 (6.1)	60.3 (5.3)
D.b.h. (in.)	15.1 (0.7)	13.2 (0.4)
Total height (ft.)	80.5 (5.0)	64.4 (3.7)
Age (yrs.)	37.0 (2.4)	34.3 (3.5)
Weight (lbs.)	2,903.2 (518.0)	2,759.5 (429.8)

[1] Graduate Student, Assistant Professor, and Research Professor, respectively, Arkansas Forest Resources Center, University of Arkansas School of Forest Resources, Monticello, AR 71656. Phone: 870–460–1993; fax: 870–460–1092; e-mail: doruska@uamont.edu.

Table 2.—*Distribution of sample trees by d.b.h. and total height class*

D.b.h. class	Total height by 5-ft class												
	45	50	55	60	65	70	75	80	85	90	95	100	Total
10	1	1	1		1	1	2						7
12	1		3	3	2	5	5	6	1				26
14				5	5	5	17	10	7	2	2		53
16				1	2	4	9	4	4	3	2	2	31
18						1	6	5	4	3	2	2	23
20							2		3	2	2		9
22							1					1	2
24													0
26									1			1	2
28										1			1
30								1					1
Total	2	1	4	9	10	16	43	26	20	10	10	4	155

(inches) at breast height were taken on each of the study trees using a diameter tape, clinometer, and a bark gauge. Table 2 shows the distribution of all study trees based on d.b.h. and total height. After all measurements were taken, each tree was numbered and marked with paint for identification on the log deck.

Soon after the first visit to each stand, logging contractors from Georgia Pacific Corp. harvested the trees. Each study tree was felled and brought to a designated log deck where it was delimbed and topped at approximately 4 inches dib. Within 2 days of felling, the felled trees were measured and weighed. Measurements taken on each of the felled study trees include length (feet) to a 4-inch dib top, dib at both ends, age, and heartwood diameter at both ends. Each study tree was then weighed using a digital scale, loader, chains, and tongs. Then the trees were bucked into merchantable lengths to satisfy Georgia Pacific's saw log specifications (17, 26, or 35 feet with a minimum top diameter of 8-inches inside bark). The same measurements that were taken on the felled study trees, including weight, were then taken on each of the merchantable saw logs. Each measurement taken (before and after harvest) was used later as a potential independent variable in creating regression equations.

Analysis

Data from 40 trees were set aside as a validation data set and data from the remaining 115 trees were used in building the regression models. Typical regression diagnostics were used in comparing and selecting the best equation forms. Indicator variables were used to determine if equation parameters varied significantly between winter and spring. The validation data set served as an additional diagnostic for comparing equations. The equations that best predicted weight to a 4-inch dib top and weight to an 8-inch dib top were chosen based on the diagnostics and indicator variable significance.

Model Comparisons

The final weight equations were compared with three published equations developed in the Southeast:

1. Newbold *et al.* (2001),
2. Clark and Saucier (1990), and
3. Baldwin (1987).

Each of these models was applied to the validation data set and the residuals were used to compare models. The models developed, which were created using only the regression data set, were also applied to the validation data.

Results and Discussion

Predicting Weight to a 4-inch DIB Top

Natural log transformations were needed to assure normality of errors. The best independent variables for predicting outside-bark green bole weight to a 4-inch dib top were d.b.h. and total tree height. The final 4-inch dib weight equation developed from the regression data set was:

$$\ln(\hat{W}t_i) = -0.1341 + 2.0178\ln(D_i) + 0.5726\ln(H_i) \qquad (1)$$

Where: $\hat{W}t_i$ = Predicted green weight (lbs.) outside bark to a 4-inch dib top for tree i,

D_i= d.b.h. (in.) for tree i, and

H_i= Total tree height (ft) for tree i.

All parameters were significant at the 0.05 α-level. The R^2 for equation (1) was 0.95 and the mean absolute residual was 236.34 pounds. The indicator variables were not significant for the intercept (p-value = 0.2969), $\ln(D_i)$ partial slope (p-value = 0.3467), or $\ln(H_i)$ partial slope (p-value = 0.2925), indicating that there was not enough evidence to conclude a significant difference in 4-inch dib weight equation parameters between winter and spring.

Table 3 compares the residuals obtained by applying each equation to the validation data set. It appears that all the equations overpredicted weight to a 4-inch dib top on average. According to the standard deviation and the mean absolute residual, which indicate on average how far off the regression line the actual values are, equation (1) appears to be best at predicting tree weight in south Arkansas. It was expected that our equation would be better at predicting weight of trees in validation data acquired in this study. However, this was the most objective means of comparison available.

Figure 1 depicts the results when the four equations were applied to the validation data set. There is little difference in predicted weights between any of the equations. The variation in predicted weights between models supports the idea that specific gravity varies somewhat by geographic location. As figure 1 shows, the equation developed by Clark and Saucier (1990) consistently estimates the lowest weight relative to the other equations. This is followed (in order from lightest to heaviest) by equation (1), Newbold *et al.* (2001) and Baldwin (1987). This corresponds to the variation in loblolly pine specific gravity by location found by the USDA (1965). Loblolly pine specific gravity in the Southeastern United States tends to increase from east to west and from north to south. This illustrates the impacts of using weight equations in this region that were developed in other geographic locations.

Figure 1.—*Comparison of 4-inch dib top equations using the validation data set.*

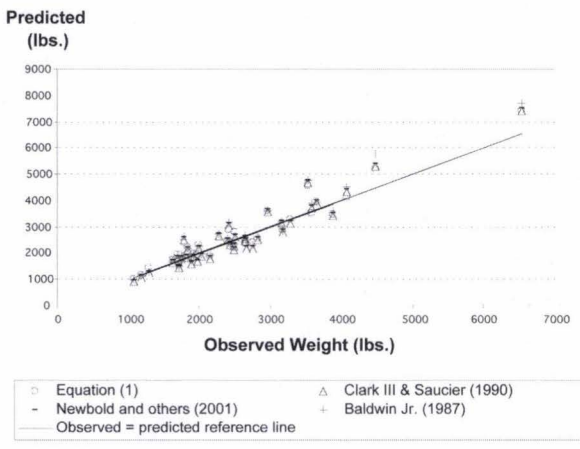


Table 3.—*Comparison of 4-inch dib top weight equations*

Attribute (lbs.)	Equation (1)	Clark and Saucier (1990)	Newbold *et al.* (2001)	Baldwin (1987)
Mean residual	-87	-25	-118	-146
Standard deviation	345	401	404	429
Mean abs. residual	253	300	299	309
Max. residual	430	583	492	454
Min. residual	-1,066	-1,139	-1,240	-1,275

Predicting Weight to an 8-inch DIB Top

As with the 4-inch dib top equation, the natural log transformation was needed to assure normality of errors. The best independent variables for predicting outside bark green merchantable saw log weight to an 8-inch top were d.b.h. and the number of 17-foot logs in the tree. The final 8-inch dib top equation built from the regression data set was:

$$\ln(\hat{W}t_i) = 1.5810 + 1.9772 \ln(D_i) + 0.8174 \ln(Logs_i) \quad (2)$$

Where: $\hat{W}t_i$= Predicted green weight (lbs.) outside bark to an 8-inch dib top for tree i,

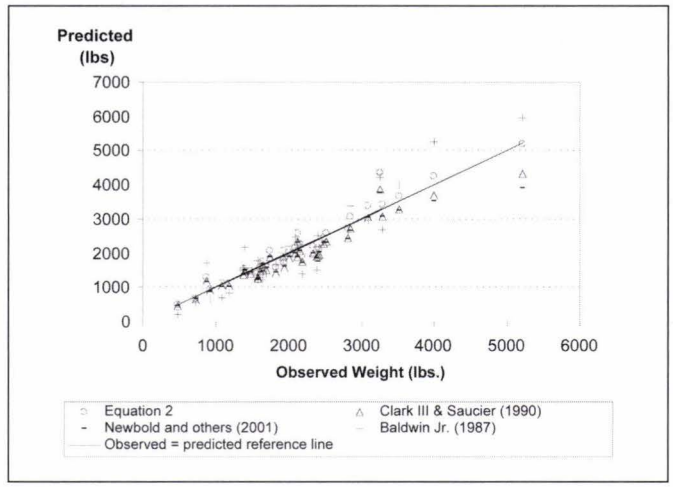
D_i= d.b.h. (in) of tree i, and

$Logs_i$= Number of 17-foot logs in tree i.

All parameters were significant at the 0.05 α-level. The R^2 for equation (2) was 0.97 and the mean absolute residual was 139.79 pounds. The indicator variables were not significant for the intercept (p-value = 0.2851), $\ln(D_i)$ partial slope (p-value = 0.2428), or $\ln(Logs_i)$ partial slope (p-value = 0.2965), indicating that there was not enough evidence to conclude a significant difference in 8-inch dib top weight equation parameters between seasons.

The residual summary in table 4 shows that on average, equation (2) slightly overpredicts weight to an 8-inch dib top, whereas the others underpredict the weight. The residuals show that there are more differences between 8-inch dib top equations than between 4-inch dib top equations. The differences between predicted and actual weight for each observation in the validation data set can be seen in figure 2. The variation in the weights predicted by the equations clearly increases as actual tree weight increases. According to table 4, equation (2) seems to be most similar to the equation developed by Clark and Saucier (1990).

Figure 2.—*Comparison of 8-inch dib top equations using the validation data set.*



Discussion and Conclusion

The best independent variables for predicting green bole weight to a 4-inch dib top were d.b.h. and total tree height. D.b.h. and the number of logs explained the most variation in green bole weight to an 8-inch dib top. There were no significant differences in 4-inch or 8-inch dib top equation parameters between winter and spring. It is important to note that the eight stands used in this study were extremely variable in site characteristics (moisture, soil type, etc.); therefore, confounding effects obviously exist. The site variability probably causes as much variation within seasons as between seasons. A more appropriate way to test seasonality would be to visit the same stands during different seasons. However, this was not possible for this study. Therefore, the only sound conclusion to be drawn here regarding seasonality is that there were no differences in weight equation parameters between winter and spring when averaging across all types of sites used in this study. Our equations were

Table 4.—*Comparison of 8-inch dib top weight equations*

Attribute (lbs.)	Equation (1)	Clark and Saucier (1990)	Newbold *et al.* (2001)	Baldwin (1987)
Mean residual	-41	163	188	18
Standard deviation	254	248	289	478
Mean abs. residual	171	229	249	380
Max. residual	316	880	1,294	903
Min. residual	-1,119	-595	-501	-1,276

most similar to the equations created by Clark and Saucier (1990), which were developed in Georgia and the surrounding States. The 4-inch and 8-inch dib top equations presented in this study should be sufficient for estimating outside-bark green bole weight of loblolly pine sawtimber in south Arkansas.

Acknowledgments

Funding for this project was provided by the Arkansas Forest Resources Center. In-kind services were provided by Georgia Pacific Corporation, Plum Creek Timber Company, and logging contractors L.D. Long and Rob Jones. This paper is published with the approval of the Director, Arkansas Agricultural Experiment Station.

Literature Cited

Baldwin, V. Clark, Jr. 1987. Green and dry weight equations for above ground components of planted loblolly pine trees in the West Gulf Region. Southern Journal of Applied Forestry. 11(4): 212–218.

Clark, Alexander, III; Saucier, Joseph R. 1990. Tables for estimating total tree weights, stem weights, and volumes of planted and natural southern pines in the Southeast. Georgia For. Res. Pap. 79. Georgia Forestry Commission, Research Division. 23 p.

Newbold, Ray A; Baldwin, V. Clark, Jr.; Hill, Gary. 2001. Weight and volume determination for planted loblolly pine in North Louisiana. Res. Pap. SRS-26. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 22 p.

U.S. Department of Agriculture. 1965. Status report; Southern wood density survey. Res. Pap. FPL-26. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 38 p.

Self-referencing Taper Curves for Loblolly Pine

Mike Strub[1], Chris Cieszewski[2], and David Hyink[3]

Abstract.—We compare the traditional fitting of relative diameter over relative height with methods based on self-referencing functions and stochastic parameter estimation using data collected by the Virginia Polytechnic Institute and State University Growth and Yield Cooperative. Two sets of self-referencing equations assume known diameter at 4.5 feet inside (dib) and outside (dob) bark; and one set assumes known only dob at 4.5 feet. A fourth degree polynomial in one minus relative height describes taper. The proposed methods improved the error sum of squares by up to 47 percent over the traditional method.

We demonstrate here the application of self-referencing curves to taper equations. Individual tree taper is commonly modeled by predicting relative diameter (diameter at a reference height divided by diameter at breast height) as a function of relative height (reference height divided by total height). Many of the models proposed in the past have not resulted in estimated diameter outside bark at 4.5 feet equal to diameter at breast height. Valentine and Gregoire (2001) proposed a model that forces the estimated diameter outside bark to be equal to diameter at breast height. Goelz and Burk (1996) and Cieszewski *et al.* (2000) argue that such constraints result in parameter estimates describing biased curve shapes. Our study proposes parameter estimation techniques that avoid such a bias while using equations that estimate diameter outside bark at 4.5 feet equal to diameter at breast height.

Data

Data for this study were collected by the Virginia Polytechnic Institute and State University Growth and Yield Cooperative. Two trees were felled at initial (1980-1982 dormant seasons) and second thinning and in buffers of unthinned control plots (1992-1994 dormant seasons) in a thinning study established at 186 locations across the natural range of loblolly pine. Disks were cut from felled trees beginning at the stump and every 4 feet to approximately a 2-inch top. Diameter inside and outside bark were measured for each disk. Diameter at breast height and total height were also measured on each tree.

Methods

To model tree taper we used a fourth degree polynomial in relative diameter versus relative height having two inflection points, which is a desirable characteristic of taper equations. We further assumed that taper curves are anamorphic in nature.

This results in the following model:

$$\frac{d}{dbh} = a_1(1-h/ht) + a_2(1-h/ht)^2 + a_3(1-h/ht)^3 + a_4(1-h/ht)^4,$$

where h = reference height, d = diameter inside or outside bark at the reference height, d.b.h. = diameter at breast height, ht = total height, d/d.b.h. = relative diameter, and h/ht = relative height.

Common practice has been to fit this equation to all the data pairs, or to multiply both sides of the equation by d.b.h. to obtain an equation in diameter inside or outside bark that can be fit to the data:

$$d = dbh \cdot a_1(1-h/ht) + a_2(1-h/ht)^2 + a_3(1-h/ht)^3 + a_4(1-h/ht) \quad (1)$$

[1] Forest Biometrician, Weyerhaeuser Company, P.O. Box 1060, Hot Springs, AR 71902. Phone: 501–624–8504; fax: 501–624–8505; e-mail: mike.strub@weyerhaeuser.com.

[2] Assistant Professor of Fiber Supply Assessment, The University of Georgia, Warnell School of Forest Resources, Athens, GA 30602–2152. Phone: 706–542–8169; fax: 706–542–8356; e-mail: biomat@smokey.forestry.uga.edu.

[3] Senior Scientific Advisor, Weyerhaeuser Company, P.O. Box 9777, Federal Way, WA 98063–9777. Phone: 253–924–6315; fax: 253–924–6736; e-mail: dave.hyink@weyerhaeuser.com.

This is analogous to fitting a guide curve to dominant height and age pairs when developing site index curves. Hereafter we refer to fitting equation (1) to the data as fitting a guide curve. An alternate approach is to obtain a self-referencing curve according to the generalized algebraic difference approach developed by Cieszewski and Bailey (2000). We introduce into the equation an unobservable variable, X, that varies from tree to tree.

$$\frac{d}{dbh} = X\left[a_1(1-h/ht) + a_2(1-h/ht)^2 + a_3(1-h/ht)^3 + a_4(1-h/ht)^4\right]$$

Multiplying both sides of the equation by d.b.h. results in:

$$d = dbh \bullet X[a_1(1 - h/ht) + a_2(1 - h/ht)^2 + a_3(1 - h/ht)^3 + a_4(1 - h/ht)^4]$$

However, this model is overparameterized. The substitution of Z = a_4 • dbh • x results in:

$$d = Z\left[b_1(1-h/ht) + b_2(1-h/ht)^2 + b_3(1-h/ht)^3 + (1-h/ht)^4\right] \quad (2)$$

where $b_1=a_1/a_4$, $b_2=a_2/a_4$, and $b_3=a_3/a_4$, and Z is estimated for each tree by minimizing the error sum of squares for each tree:

$$SSe_j = \sum_{i=1}^{Nj}\left(d_{ij} - Y_j f_{ij}\right)^2 \quad (3)$$

where:

SSe_j is the error sum of squares for the jth tree

N_j is the number of diameter measurements for the jth tree

d_{ij} is the ith diameter measurement for the jth tree

Z_j is a level parameter for the jth tree

$$f_{ij} = b_1(1-h_{ij}/ht_j) + b_2(1-h_{ij}/ht_j)^2 + b_3(1-h_{ij}/ht_j)^3 + (1-h_{ij}/ht_j)^4$$

h_{ij} is the height of the ith diameter measurement for the jth tree

ht_j is the total height of the jth tree

Taking the first derivative of SSe_j with respect to Z_j, setting it equal to zero and solving for Z_j results in the least squares estimate of Z_j:

$$Z_j = \frac{\sum_{i=1}^{N_j} d_{ij} f_{ij}}{\sum_{i=1}^{N_j} f_{ij}^2} \quad (4)$$

These estimates can be used in the SAS NLIN procedure to obtain least squares estimates of the global parameters b_2, b_3, and b_4. The value of Z_j can be calculated given the global parameters and retained for subsequent data points from the same tree. Details of this method are described in Strub and Cieszewski (2002) for the more general case where Z_j cannot be analytically solved for in a closed form. Note that this method is not valid if a_4 is zero. However, if a_4 is zero, the model would have only one inflection point and would not be appropriate for a taper function. Substituting equation (3) into equation (2) results in minimizing the error sum of squares.

(5)

$$SSe_j = \sum_{i=1}^{Nj}\left(d_{ij} - \frac{\sum_{i=1}^{N_j} d_{ij} f_{ij}}{\sum_{i=1}^{N_j} f_{ij}^2} f_{ij}\right)^2 \text{; and } \quad SSe = \sum_{j=1}^{M}\sum_{i=1}^{Nj}\left(d_{ij} - \frac{\sum_{i=1}^{N_j} d_{ij} f_{ij}}{\sum_{i=1}^{N_j} f_{ij}^2} f_{ij}\right)^2$$

Numerical techniques must be used to estimate the global parameters b_2, b_3, and b_4 by minimizing this error sum of squares. Separate parameters can be obtained for diameter inside and diameter outside bark equations. Application of these equations requires an estimate of both diameter inside and outside bark at some index height, usually 4.5 feet, just as application of site index curves requires an estimate of dominant height at an index age (Cieszewski *et al.* 2000). Given d.b.h., the following equation describes the outside bark taper profile.

(6)

$$d = \frac{dbh\left[b_1(1-h/ht) + b_2(1-h/ht)^2 + b_3(1-h/ht)^3 + (1-h/ht)^4\right]}{b_1(1-4.5/ht) + b_2(1-4.5/ht)^2 + b_3(1-4.5/ht)^3 + (1-4.5/ht)^4}$$

If bark thickness is measured at d.b.h., diameter inside bark at 4.5 feet can be substituted for d.b.h. in equation (6) to obtain the inside bark taper profile. Often only diameter outside bark at 4.5 feet (diameter at breast height) is known. A system of equations that predicts both diameter inside and outside bark from only diameter at breast height can be developed by modeling Z in the inside bark equation as a linear function of Z in the outside bark equation (see justification for this assertion in the Results section). Implementing this maneuver results in the following definition of the error sum of squares for the jth tree when both inside and outside bark measurements are given the same weight.

$$SSe_j = \sum_{i=1}^{Nj} \left(dib_{ij} - Z_j f_{ij}\right)^2 + \sum_{i=1}^{Nj} \left(dob_{ij} - (a_0 + a_1 Z_j) g_{ij}\right)^2 \quad (7)$$

Where:

$f_{ij} = bib_1 (1 - h_{ij} / ht_j) + bib_2 (1 - h_{ij} / ht_j)^2 + bib_3 (1 - h_{ij} / ht_j)^3 + (1 - h_{ij} / ht_j)^4$

$g_{ij} = bob_1 (1 - h_{ij} / ht_j) + bob_2 (1 - h_{ij} / ht_j)^2 + bob_3 (h_{ij} / ht_j)^3 + (1 - h_{ij} / ht)^4$ bib_1, bib_2, bib_3, bob_1, bob_2, bob_3, a_0, a_1 are global parameters.

Minimizing the error sum of squares with respect to Z_j

$$Z_j = \frac{\sum_{i=1}^{N_j} dob_{ij} g_{ij} + a_1 \sum_{i=1}^{N_j} dib_{ij} f_{ij} + a_0 a_1 \sum_{i=1}^{N_j} f_{ij}^2}{\sum_{i=1}^{N_j} g_{ij}^2 + a_1^2 \sum_{i=1}^{N_j} f_{ij}^2}$$

The remaining eight global parameters (bib_1, bib_2, bib_3, bob_1, bob_2, bob_3, a_0, a_1) can be estimated with non-linear regression. Once the parameters are estimated, a single Z value can be calculated from d.b.h. by using equation (8):

$$Z = \frac{dbh}{bob_1(1-4.5/ht) + bob_2(1-4.5/ht)^2 + bob_3(1-4.5/ht)^3 + (1-4.5/ht)^4}$$

The outside-bark taper profile can be determined from equation (9) and the inside-bark equation can be estimated from equation (10):

$$dob = Z\left[bob_1(1-h/ht) + bob_2(1-h/ht)^2 + bob_3(1-h/ht)^3 + (1-h/ht)^4\right]$$

$$dib = (a_0 + a_1 Z)\left[bib_1(1-h/ht) + bib_2(1-h/ht)^2 + bib_3(1-h/ht)^3 + (1-h/ht)^4\right]$$

Results

Equation (1) was fit to the data resulting in a guide curve that is similar to the traditional methods of modeling tree taper. The model is shown for both diameter outside bark in black and diameter inside bark in gray in figure 1a. Parameter estimates are given in table 1. Given d.b.h. and total height, diameter outside bark at 4.5 feet or d.b.h. can be estimated from the outside bark guide curve. The difference between the estimated d.b.h. and the input d.b.h. is shown in figure 1b for a range of input d.b.h.'s and total heights. Note the large discrepancy between input d.b.h. and estimated d.b.h. for many reasonable d.b.h. and total height pairs. This difference suggests that methodology for ensuring the consistency of input d.b.h. and estimated d.b.h. is well grounded.

Figure 1c shows relative diameter outside bark versus relative height for each data observation. Most of the data are plotted with gray circles representing each data point. Four selected trees were plotted as black diamonds, crosses, triangles, or squares. Note that these four trees tend to lie at the top, middle, or bottom of the data points. This suggests that a system of curves analogous to site index curves could better approximate tree taper. A single guide curve fit through the middle of the data range will not adequately describe trees with taper profiles like the black diamonds and squares.

Equation (2) was fit to both inside and outside bark diameter. Parameter estimates are given in table 2. The taper profile for selected total heights is shown for diameter outside bark in

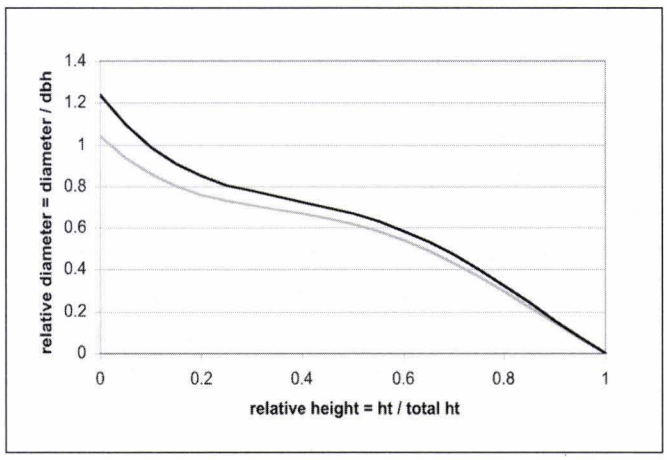
Table 1.—*Parameter estimates for the guide curve, equation 1*

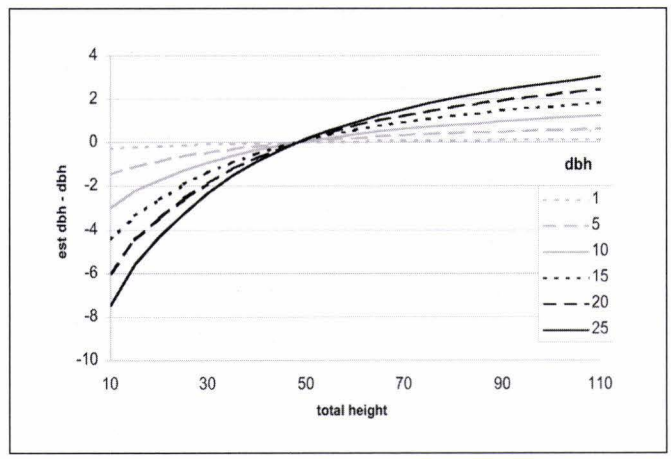
Model	Parameter	Estimate
Diameter outside bark	a_1	1.4195
	a_2	2.1537
	a_3	-6.9315
	a_4	4.5998
Diameter inside bark	a_1	1.2879
	a_2	1.9968
	a_3	-6.1580
	a_4	3.9148

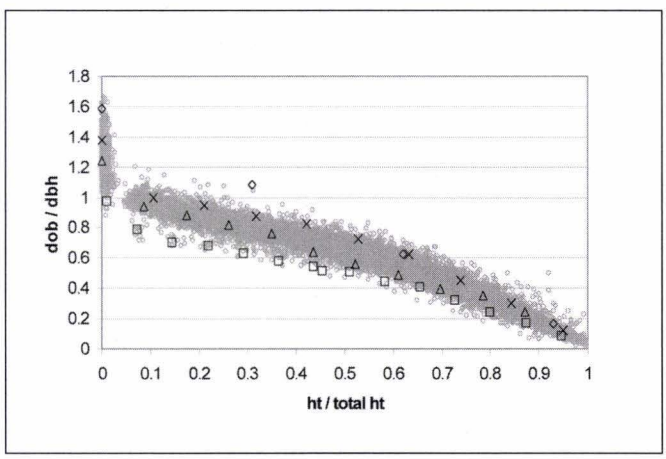
Table 2.—*Parameter estimates for the self-referencing curve, equation 2*

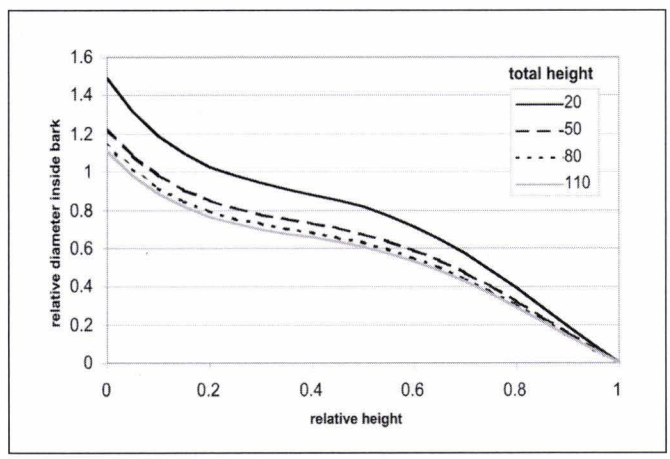
Model	Parameter	Estimate
Diameter outside bark	b_1	0.314256
	b_2	0.462600
	b_3	-1.504593
Diameter inside bark	b_1	0.330808
	b_2	0.507229
	b_3	-1.571667

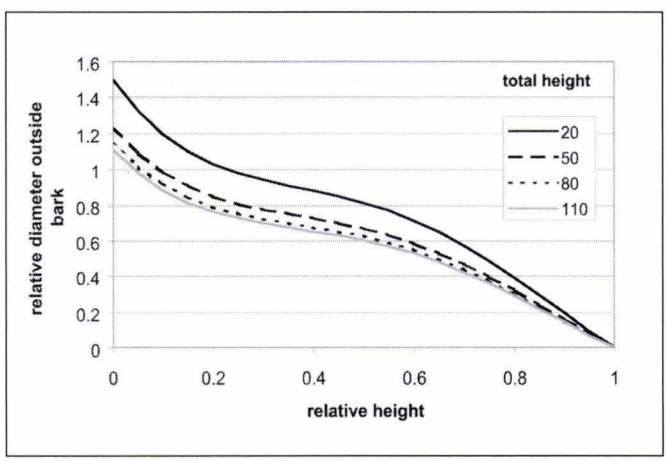
Figure 1.—(a) *Relative diameter inside and outside bark versus relative height for the guide curve, equation (1); (b) error in d.b.h. estimate obtained from the outside bark guide curve; (c) observed relative diameter outside bark versus relative height is shown for the majority of the data in gray circles (four selected trees were plotted as black diamonds, crosses, triangles, or squares); (d) relative diameter outside bark versus relative height for selected total heights and the self-referencing curve, equation (2); (e) relative diameter inside bark versus relative height for selected total heights and the self-referencing curve, equation (2); and (f) Zdob versus Zdib and the linear fit (the Zdob = Zdib line is shown in gray).*











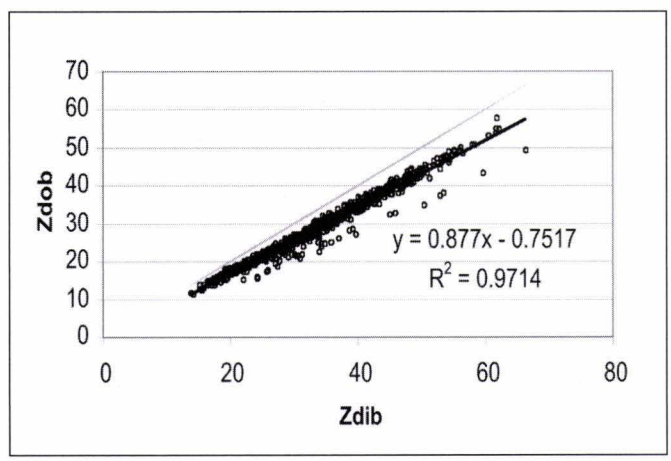


figure 1d and diameter inside bark in figure 1e. Note that the taper profile changes dramatically as total height changes. Fitting the outside bark equation resulted in an estimate of Z_j, Zdob, for each tree, and fitting the inside bark equation resulted in an estimate of Z_j, Zdib, for each tree. Applying these taper equations (eq. (6)) requires estimating both Zdob and Zdib. Obtaining these estimates requires knowledge of both diameter inside and outside bark at some reference point usually breast height or 4.5 feet. An alternate approach was developed that requires only d.b.h. Figure 1f shows the strong linear relationship between Zdob and Zdib. This suggests that error sum of squares described in equation (7) could be minimized to obtain inside and outside bark taper estimates based on equation (9) and inside bark estimates based on equation (10). A common Z is estimated from d.b.h. using equation (8). Table 3 gives parameter estimates for these equations.

Discussion and Conclusion

The base-age-invariant parameter estimation (Bailey and Clutter 1974) used in this research can be successfully fit with sufficient data and fairly straightforward programming using SAS (SAS Institute 1990). The data must consist of repeated measures, which are pooled cross-sectional and longitudinal data. The number of measurements in each series has to be at least two, and the number of series must be greater than the number of global parameters considered in the model. The proposed method does not violate regression principles.

Table 3.—*Parameter estimates for the self-referencing curves, equations 8, 9, and 10*

Model	Parameter	Estimate
Diameter outside bark	bob_1	0.311990
	bob_2	0.464576
	bob_3	-1.505517
Diameter inside bark	a_0	-0.009822
	a_1	0.850341
	bib_1	0.333494
	bib_2	0.505446
	bib_3	-1.571255

Table 4.—*Sum of squared errors for the guide curve approach and self-referencing equations*

Model	Guide curve	Self-referencing equations
Diameter inside bark	2372.8	1098.0
Diameter outside bark	2436.7	1456.8
Total inside bark and outside bark	4809.5	2554.8
Zib=ao+a1*Zob		2912.3

Error sum of squares are listed in table 4 for the guide curve approach, the self-referencing curves when both diameter inside and outside bark are known at breast height, and the self referencing curves when only d.b.h. is known. The total error sum of squares for both inside and outside bark guide curves was reduced by 47 percent when self-referencing curves based on separate Zdob and Zdib were used to describe tree taper. The total error sum of squares for both inside and outside bark guide curves was reduced by 39 percent when self-referencing curves based on a common Z were used to describe tree taper. These curves have the additional advantage of returning estimated diameter outside bark at 4.5 feet equal to d.b.h. The senior author has applied the same techniques to the more complex equations of Max and Burkhart (1976) with equal success, which demonstrates the broad utility of self-referencing curves.

The proposed methods are suitable for fitting applications with other dependent variables such as per acre basal area, survival, and yield.

Acknowledgment

The authors thank the Virginia Polytechnic Institute Growth and Yield Cooperative for supplying the taper data for this study.

Literature Cited

Bailey, R.L.; Clutter, J.L. 1974. Base-age invariant polymorphic site curves. Forest Science. 20(2): 155–159.

Cieszewski, C.J.; Bailey, R.L. 2000. Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes. Forest Science. 46(1): 116–126.

Cieszewski, C.J.; Harrison, M.W.; Martin, S.W. 2000. Examples of practical methods for unbiased parameter estimation in self-referencing functions. In: Cieszewski, C.J., ed. Proceedings, 1st International conference on measurements and quantitative methods and management; 1999 November 17–18; Jekyll Island, GA.

Goelz, J.C.G.; Burk, T.E. 1996. Measurement error causes bias in site index equations. Canadian Journal of Forest Research. 26: 1586–1593.

Max, T.A.; Burkhart, H.E. 1976. Segmented polynomial regression applied to taper equations. Forest Science. 22(3): 283–289.

SAS Institute. 1990. SAS guide to macro processing, version 6, 2d ed. Cary, NC: SAS Institute, Inc.

Strub, M.R.; Cieszewski, C.J. 2002. Fitting global site index parameters when plot or tree site index is treated as a local nuisance parameter. In: Proceedings, symposium on statistics and information technology in forestry; 2002 September 8–12; Blacksburg, VA.

Valentine, H.T.; Gregoire, T.G. 2001. A switching model of bole taper. Canadian Journal of Forest Research. 31: 1400–1409.

D.B.H. and Survival Analysis: A New Methodology for Assessing Forest Inventory Mortality

Christopher W. Woodall[1], Patricia L. Grambsch[2], and William Thomas[2]

Abstract.—Tree mortality has typically been assessed in Forest Inventory and Analysis (FIA) studies through summaries of mortality by location, species, and causal agents. Although these methods have historically been used for most of FIA's tree mortality analyses, they are inadequate for robust assessment of mortality trends and dynamics. To offer a new method of analyzing tree mortality in forest inventories, we used survival analysis techniques to estimate survival and hazard functions for FIA periodic inventories in Minnesota. The study's method for applying survival analysis techniques to FIA inventories successfully estimates survivor and hazard models. Classifying trees into classes of d.b.h. and d.b.h. growth may facilitate applying of survival analysis techniques by providing a surrogate for tree ages and vigor. Applying survival analysis techniques to forest inventories allows FIA inventory analysts to test tree mortality hypotheses and summarize regional tree mortality trends, and affords a solid foundation for development of individual tree mortality models.

Tree mortality in forest inventories has traditionally been assessed using simple summary statistics. Mortality information in Forest Inventory and Analysis (FIA) State reports has typically included losses in timber volume due to mortality, summaries of mortality causal agents, locations of dead trees, and mortality trends by species (Leatherberry *et al.* 1995). More in-depth mortality analysis has been facilitated only through development of logistic regression models of individual tree mortality, a technique that is cumbersome and inadequate for large forest inventories (Eid and Tulius 2001). Current forest mortality analytical techniques lack methods for incorporating the time-dependent nature of tree mortality, testing hypotheses, censoring observations, and conducting tests for effects of covariates (i.e., stand basal area and crown ratio). Given the past diseases and epidemics that have greatly altered North American forest ecosystems (e.g., chestnut blight [*Cryphonectria parasitica*] and Dutch elm disease [*Ceratocystis ulmi*]) and the threats of future forest health hazards, novel and statistically robust techniques for assessing forest mortality would greatly benefit forest inventory analysts.

Analytical methods developed by the medical sciences, collectively termed survival analysis, may provide the basis for developing new forest mortality analytical techniques. Survival analysis is usually defined as a class of statistical methods for studying the occurrence and timing of events, such as death (Allison 1995, Collett 1994). Waters (1969) first proposed using survival analysis to address forest mortality, but such applications have been restricted to forest inventories in even-aged forest plantations (Morse and Kulman 1984, Wykoff and Clark 2000) due to the inherent lack of detailed time and age information for larger-scale inventories (Flewelling and Monserud 2002). Given the current dearth of forest inventory mortality analysis techniques, a re-examination of the basics of survival analysis in the context of the FIA inventories is warranted and may provide a novel mortality analysis methodology.

The primary goal of our study was to estimate and interpret the central functions of survival analysis (survivor and hazard functions) for an FIA inventory in the State of Minnesota. Specific objectives included:

1. To use d.b.h. and d.b.h. growth (Δd.b.h.) in applying survival analyses techniques to forest inventories.
2. To determine if survivor/hazard functions can represent actual mortality trends in a manner practical for ecological interpretation.

[1] Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108. Phone: 651–649–5141; e-mail: cwoodall@fs.fed.us.
[2] Associate Professors, School of Public Health, University of Minnesota, Minneapolis, MN 55455.

Forest Survival Analysis Supposition

Time to an event is the defining component of survival methods. Hence, the major limitation often cited for the limited application of survival analysis to forest inventories is the lack of specific tree ages and the censoring of tree mortality (Flewelling and Monserud 2002). However, knowledge of age is not necessary for implementing survival analyses (Allison 1995). Any measurement unit that indicates changes in an individual's status between remeasurements may replace the traditional survival analysis variables of age and time. For forest inventories that remeasure trees at regular intervals, d.b.h. and Δd.b.h. (time 2 d.b.h. - time 1 d.b.h.) may assign individual trees to cells within a matrix of tree size and vigor. Whereas medical studies may determine survivor functions for demographic cohorts across calendar years, forest inventory survival functions may be determined for d.b.h. classes across vigor classes. The survivor function $S(t)$ is defined at a time t as the probability that the time to the event is greater than or equal to t (Collett 1994). In this study, the "clock" starts at the first forest inventory, when a subject begins to be "at risk" for the event or begins to be monitored for the event. Stating this in terms of d.b.h., the clock is Δd.b.h. (the increase in d.b.h. from initial survey). Our survival function $S(\Delta d.b.h.)$ gives the probability that a tree will die after it has grown by at least $\Delta d.b.h. = k\ cm$. For example, $S(4\ cm)$ estimates the proportion of the population of trees within the same d.b.h. class that will survive to increase their d.b.h. by 4 cm. Related to the survival function is the hazard function, $h(t)$. The hazard function gives the probability of an event occurring at time t given that the subject has survived up to t. In terms of d.b.h., $h(\Delta d.b.h.)$ gives the probability that a tree that has survived and grown k cm will die at that size. Given the robust and established analyses of the survival modeling community, the individual tree variables of d.b.h. and Δd.b.h. may allow applying survival analysis to forest inventories, thereby providing a novel method of assessing forest mortality dynamics.

Methods

Survival analysis was conducted using data from the 1977 and 1990 periodic FIA inventories for the State of Minnesota (table 1). Individual trees (observations) were included that met the following criteria: alive at time 1 and observed as either dead or alive at time 2, d.b.h. ≥ 13.0 cm (rounded up, minimum d.b.h. for subplot trees as defined by FIA program), and no human-caused mortality. Additionally, to streamline the large data sets, only the most common species representing a wide range of growth habits and suffering from a variety of damage agents were selected for each State (table 1). Δd.b.h. was calculated as the difference in d.b.h. between time 1 and time 2. If a tree was dead at time 2, its d.b.h. was equal to the d.b.h. at time 2 or the d.b.h. at time 1, whichever was larger. Since a tree's d.b.h. may shrink following death, an estimate of the maximum d.b.h. the tree attained before death would better benefit survival analysis than an estimate of a decaying bole diameter.

All data set trees were grouped both by initial d.b.h. (10-cm d.b.h. classes) and Δd.b.h. (4-cm classes). PROC LIFETEST (SAS 1999) and its life-table estimation method were used to

Table 1.—*FIA inventory for the State of Minnesota used in survival analysis*

Species Group	Species	Number of trees
Red and jack pine	*Pinus banksiana, Pinus resinosa*	3,935
Black spruce and balsam fir	*Picea mariana, Abies balsamea*	14,972
Maples	*Acer saccharinum, Acer saccharum*	2,747
Balsam poplar	*Populus balsamifera*	4,448
Paper birch	*Betula papyrifera*	8,603
American elm	*Ulmus americana*	3,829
Aspen	*Populus tremuloides*	21,303
Red oak	*Quercus rubra*	2,962

estimate $S(\Delta d.b.h.)$ and $h(\Delta d.b.h.)$ by 10-cm d.b.h. classes for the entire data set. Additionally, $h(\Delta d.b.h.)$ was estimated for the 23.0- to 32.9-cm d.b.h. class, stratified by species groups.

Results and Discussion

A survival function $S(\Delta d.b.h.)$ was estimated for selected species in the 1977-1990 Minnesota inventories (fig. 1). The survival function was estimated separately for five initial d.b.h. classes. The survivor function displays the cumulative probability of trees surviving to the inventory remeasurement (time 2) across classes of Δd.b.h. For trees of a midsize diameter, there was approximately a 60-percent probability of mortality for trees growing less than 4 cm during the remeasurement interval. Using d.b.h. and Δd.b.h. for survival analysis application, the survivor function quantifies the stand dynamics that may cause tree mortality. The greatest tree mortality occurs in trees growing 4 cm or less during the inventory interval (13 years). The largest trees suffer greater mortality rates than smaller trees. In contrast to the survivor function, the hazard function expresses the rate of death at a specific interval midpoint (Δd.b.h. class), allowing mortality trends to be broadly assessed by d.b.h. and Δd.b.h. classes (fig. 2). Hazard functions varied both by initial d.b.h. classes and Δd.b.h. The largest trees with the smallest Δd.b.h. had the highest risk (hazard) of death, while smaller trees had lower hazards of death in the smaller classes of Δd.b.h. To examine $h(\Delta d.b.h.)$ across species groups, the hazard functions for the 23.0- to 32.9-cm d.b.h. class, stratified by species group, were determined (fig. 3). Risk of mortality was distinctly different between all species groups across all classes of Δd.b.h. American elm had the greatest hazard function across all classes of Δd.b.h., while maples had the lowest hazard function.

The survivor and hazard functions may offer robust tools for analyzing forest mortality. The survivor function displays mortality cumulatively through the diameter distribution, while the hazard function may display specific d.b.h. midpoint mortality rates. As evident from the survivor and hazard function curves for Minnesota, d.b.h. classes with divergent or atypical mortality trends may be readily identified. For those that monitor forest health across regions of the United States, the analytical ability to identify and discern differences in mortality trends is crucial. We suggest that survivor curves for "typical" mortality may assume a characteristic survivor curve form. Divergences of survivor function curves from the "typical" curve bounds for specific tree populations may help identify problems in a rapid, statistically defensible manner. For large forest inventories, hazard functions may be able to attribute mortality to causal agents,

Figure 1.—*Survival functions for time one diameter classes by delta DBH (Time 2 DBH - Time 1 DBH).*

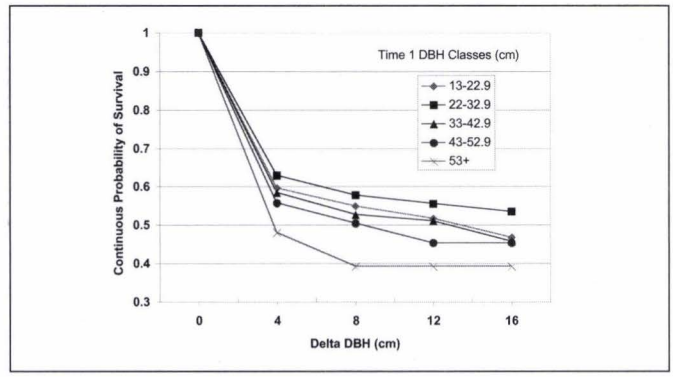


Figure 2.—*Hazard functions for time one diameter classes by delta DBH (Time 2 DBH - Time 1 DBH).*

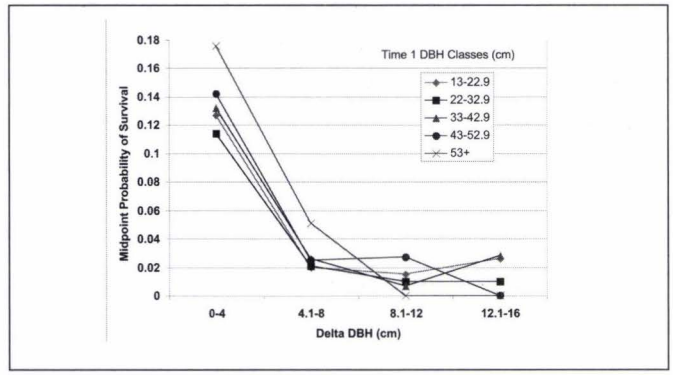
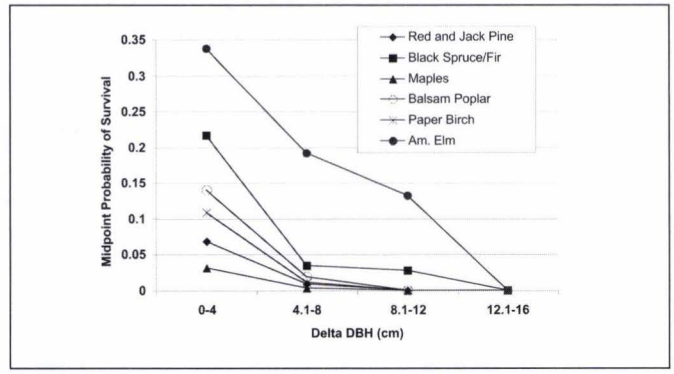


Figure 3.—*Hazard functions for time one DBH class 23-32.9 (cm) for various MN species.*



further refining forest health assessments. The hazard and survivor functions can together provide a rapid and comprehensive assessment of tree mortality for forest inventories as long as the survey interval time is approximately the same between remeasurements (FIA annual inventory remeasurement interval ≅ 5 years).

Interspecific tree mortality differences are critical to forest health assessments. Hazard functions, determined through this study's methodology, allow for comparing mortality risk rates among species and diameter classes. Although analysis using only one diameter class was presented from this study, there were obvious differences in hazard functions among species. This study's methodology may allow comparing hazard functions among species over successive inventory cycles. Detection monitoring of atypical mortality may be better facilitated through observing risks of mortality by species, d.b.h. class, and Δd.b.h. class (hazard functions).

A longitudinal unit can be any unit that measures a variable's transition from one state (i.e., class or condition) to another (Collett 1994). The greatest hurdle in applying survival analytical techniques to forest inventories is finding appropriate longitudinal units to quantify the transition of individual trees from alive to dead. If time or ages are used as longitudinal units in forest inventory analyses, a number of problems may be encountered. First, all observations are censored. The exact time of tree death is uncertain, with the inventory remeasurement date often serving as the longitudinal measure. Second, the survivor function curve is partially dependent on when and where the measurements were taken. For example, if the bulk of mortality is located in a certain area of the State that is inventoried at a discrete point in time, the resulting survival curve will be biased if time is used. Third, the age of a tree is difficult to estimate in large forest inventories. However, d.b.h. and Δd.b.h. are quantities that hypothetically increase until a tree dies. Thus, tree diameter may be used as a surrogate of age in survival analysis. Δd.b.h., although not a surrogate for time, may serve as a "stopwatch" for individual trees. At the start (time 1), the Δd.b.h. of all trees is 0. At the time of remeasurement (time 2) the "stopwatch" is stopped and trees are assigned to classes of Δd.b.h.. Time (years) may greatly relate to the survival of humans, while tree growth over intervals of time (i.e., annual diameter growth) may be a more meaningful measure in forest ecology. Using the variables of tree size and growth may allow survival analyses to be conducted on forest inventories and warrant future evaluation and possible application.

Conclusion

Forest inventory mortality analysis has predominantly been focused on logistic regression modeling at the individual tree-scale with scant data summarizations at the landscape scale. This apparent disparity in research efforts between forest ecosystem scales means few advances or technologies have been forwarded for robust analysis of forest mortality dynamics at the landscape scale. This study proposed a new approach to forest mortality assessment by combining established survival modeling techniques (survivor/hazard functions) with traditional measurements of forest stand attributes (d.b.h. distribution/diameter growth). This technique suggests a paradigm shift in forest mortality analyses and nonstandard application of survival analysis techniques. If this study's techniques withstand the test of time and peer review, a new forest mortality analysis approach may be gained that is more efficient and provides statistically defensible assessments of tree mortality for tree populations across different forest types, locations, and various damaging agents.

Literature Cited

Allison, P.D. 1995. Survival analysis using the SAS system, a practical guide. Cary, NC: SAS Institute, Inc. 300 p.

Collett, D. 1994. Modeling survival data in medical research. New York, NY: Chapman and Hall, CRC. 347 p.

Eid, T.; Tulius, E. 2001. Models for individual tree mortality in Norway. Forest Ecology and Management. 154: 69–84.

Flewelling, J.; Monserud, R.A. 2002. Comparing methods for modeling tree mortality. In: Crookston, N.L.; Havis, R.N., eds. Proceedings, 2d forest vegetation simulator conference; 2002 February 12–14; Fort Collins, CO. RMRS-P-25. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky

Leatherberry, E.C.; Spencer, J.S., Jr.; Schmidt, T.L.; Carroll, M.R. 1995. An analysis of Minnesota's fifth forest resources inventory, 1990. Resour. Bull. NC-165. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 102 p.

Morse, B.W.; Kulman, H.M. 1984. Plantation white spruce mortality: estimates based on aerial photography and analysis using a life-table format. Canadian Journal of Forest Research. 14: 195–200.

SAS Institute, Inc. 1999. SAS software, version 8. Cary, NC: SAS Institute, Inc.

Waters, W.E. 1969. Life-table approach to analysis of insect impact. Journal of Forestry. 67: 300–304.

Wykoff, P.H.; Clark, J.S. 2000. Predicting tree mortality from diameter growth: a comparison of maximum likelihood and Bayesian approaches. Canadian Journal of Forestry Research. 30: 156–167.

Analysis of Pooled FIA and Remote Sensing Data for Fiber Supply Assessment at the Warnell School of Forest Resources at the University of Georgia—Other Studies and Effective Information Dissemination

Chris J. Cieszewski[1], Michal Zasada[2, 7], Tripp Lowe[3], Bruce Borders[4], Mike Clutter[4], Richard F. Daniels[4], Robert I. Elle[5], Robert Izlar[6], and Jarek Zawadzki[2, 8]

Abstract.—We provide here a short description of the origin, current work, and future outlook of the Fiber Supply Assessment program at the D.B. Warnell School of Forest Resources, University of Georgia, whose work includes various analyses of FIA data. Since 1997, the program has intended to assist the implementation of the new Southern Annual Forest Inventory System through related data analyses. Currently its projects include basic problems in theory of equations and parameter estimation and various analyses of ground inventories and remote sensing and GIS data. We describe some of these projects and associated software, hardware, and information dissemination problems and solutions.

The Fiber Supply Assessment program (FSA) at the Daniel B. Warnell School of Forest Resources was initiated by Dean Arnett Mace, Jr. in 1997. The establishment of this program coincided with the beginning of the implementation of the Southern Annual Forest Inventory Analysis System (SAFIS) and was intended to provide the school's input into solving the various problems of timely and accurate fiber supply assessment in Georgia.

Parties interested in creation of this program included members of the forest product industry, and others. Their expectations were directed toward finding new relationships and revealing information in the new annual measurement data produced by SAFIS. The new design of the continuous annual inventory was attracting many questions about statistical accuracy, the possibility of monitoring growth, and differences between current and former periodic estimates.

Notwithstanding the above, in the beginning the FSA could not focus on the annual data analysis due to the unavailability of such data. Furthermore, there was also little pragmatic value in work on new estimators because it seemed rather futile to begin changing the barely conceived statistical design, which was so new that it was not even quite implemented yet. Thus, at the outset the FSA focused initially on theoretical studies of inventory projection equations. Subsequent efforts concentrated on building collaborative studies with other programs and exploring funding opportunities related to the general mandate of the program. Presently the program is collaborating with forest biometrics, the quantitative forest management and wood quality programs, the center for forest business, forest finance, forest economics, and a number of forest product industry partners.

Research of the Fiber Supply Assessment Program

Inventory projection equations: Forest inventory updates and projections frequently rely on a special type of equation known as a self-referencing function (Northway 1985). Such equations are applied to model such stand characteristics as basal area, volume, and different height measures. They contain two types of parameters: the global model parameters, and the initial condition parameters, which are snapshot observations of the modeled phenomena usually available through inventory measurements. These equations are used to model various characteristics that depend on unobservable variables (e.g., site productivity). Examples of the modeled characteristics in forestry

[1] Assistant Professor, [2] Postdoctoral Associate, [3] GIS Analyst, [4] Professor, [5] Research Coordinator, [6] Director of Center for Forest Business, respectively, Warnell School of Forest Resources, University of Georgia, Athens, GA.
[7] Assistant Professor at Faculty of Forestry, Warsaw Agricultural University, Poland.
[8] Assistant Professor at Environmental Engineering Department, Warsaw University of Technology, Poland.

are height (e.g., Bailey and Clutter 1974), diameter (e.g., Clutter *et al.* 1983), basal area (e.g., Pienaar and Shiver 1986), volume (e.g., Coile and Schumacher 1964), trees per unit area (e.g., Bailey *et al.* 1985), and biomass and carbon sequestration (e.g., Cieszewski *et al.* 1996). The self-referencing functions, or the inventory projection equations, are usually calibrated on pooled longitudinal and cross-sectional data and can be expressed by either dynamic or static equations.

The FSA focuses almost exclusively on the dynamic equations, also known as dynamic site equations, which are relatively scarce in forestry literature. Out of a few hundred publications on self-referencing models, only a couple of dozen use dynamic equations (Cieszewski 2001). Similarly, in spite of an abundance of literature on parameterization of explicit static equations (i.e., $Y=f(t)$), there is little literature available on derivation of these implicit dynamic equations $Y=f(t, Y)$). Our program made major contributions in this area by founding the generalization of the algebraic difference approach (Cieszewski and Bailey 2000), providing internationally examples of its applications (e.g., Cieszewski *et al.* 1999; Cieszewski, in press; Cieszewski and Nigh 2002; Cieszewski and Zasada 2002), and presenting new methods of dynamic equation derivation (e.g., Cieszewski 2001).

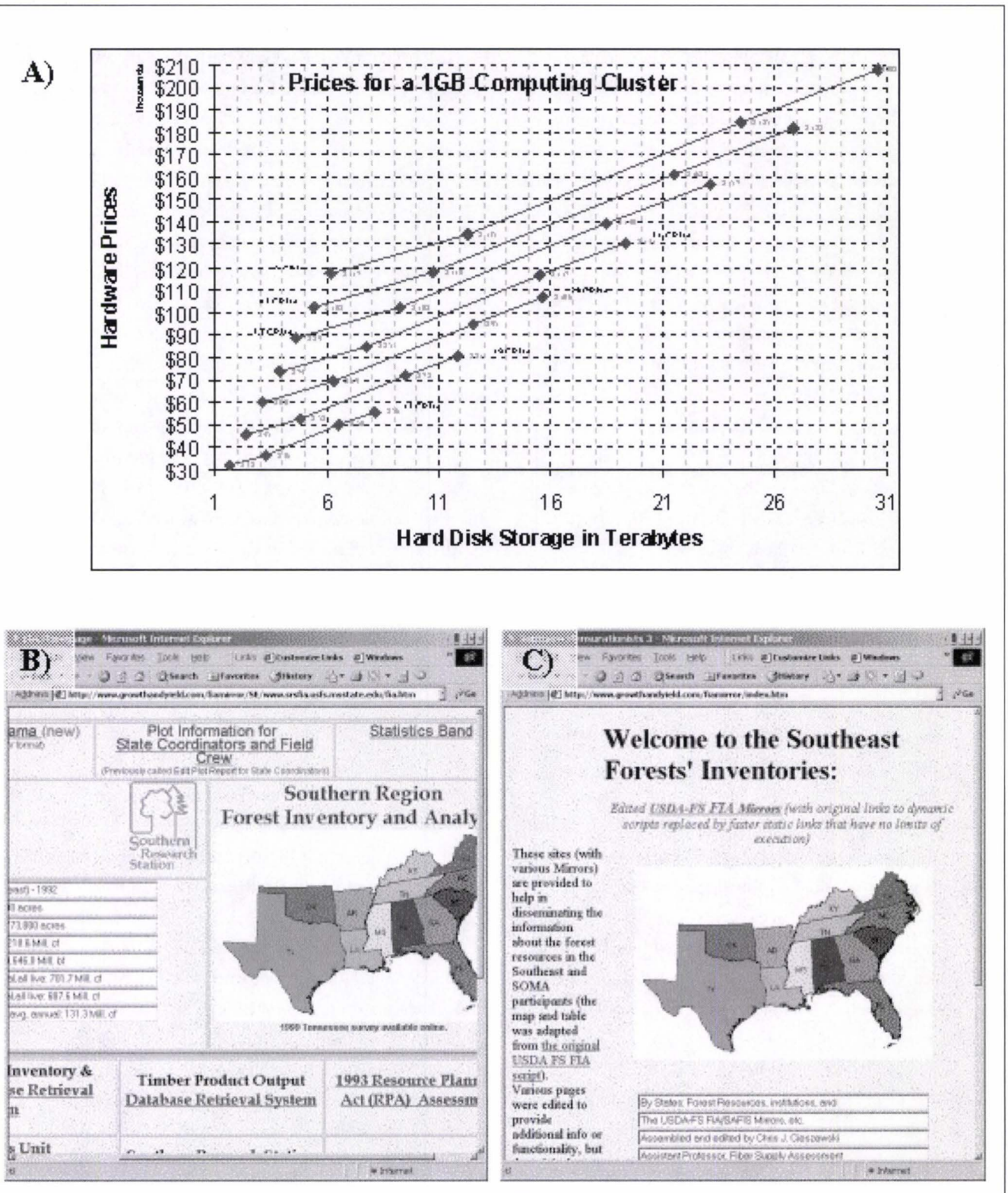
Long-term sustainability analysis: In 2000 collaboration between the Wood Quality, Forest Biometrics, and the FSA began a major research effort on long-term sustainability analysis of fiber supply in Georgia. The Dean of the School supported this effort with an initial investment of $40,000 for software and hardware necessary for initiation of spatially explicit estate simulation analysis. At the beginning of 2001 this effort created a new postdoc position and led to extensive research and analyses of various Georgia forest resource data. Subsequently, long-term research was made possible through TIP3 funding of a 3-year project with the postdoc position and graduate students. The main data for the analysis came from the FIA 1997 and 2000 surveys. Supplementary information came from various GIS and remote sensing sources including the Landsat TM 7 imagery. A preliminary approach to the analysis was described first in Cieszewski *et al.* (2001). Subsequent stages and technical details can be found in Cieszewski *et al.* (2002) and Zasada *et al.* (2002). Part of the long-term sustainability analysis project concentrates on studying impacts of intensive management practices on future wood production in Georgia. Assumptions and results for the intensive plantation management practices are described in Zasada *et al.* (2003a).

The sustainability project involves many simulations and massive numerical outputs. Reviewing the results is a considerable challenge even for the team members and much more so for the clients and public. The initial results consisted of over 1,000 graphs, maps, and tables that were organized for public access on the Web at http://www.growthandyield.com/sustain/ in a navigational tree framework with expandable/collapsible branches (fig. 1A). This presentation is fairly effective but as the number of images and text files grew it became apparent that invoking images by mouse click is not an effective way of viewing thousands of images. To solve this problem and enable more effective browsing through the massive outputs, we designed a new Web construct based on dynamic image tables (e.g., fig. 1B and 1C) being displayed automatically when a mouse cursor browses over navigation tables (on a click the display image table becomes the navigation table and a new display image table is browsed as the mouse-over display as on figure 1C). The later results from this project are being published at http://www.growthandyield.com/movesustain/ in a dynamic navigation framework optimized for fast and efficient browsing of large numbers of tables and images; however, the users need to be patient at loading time when large numbers of images are preloaded to speed up subsequent response time.

Assessment of stream and road buffers: In 2002 the USDA Forest Service agreed to sponsor a study on assessment of protective buffers such as the stream management zones and road beautifying buffers. It has been a major contribution to the research in the FSA. The technical aspects and preliminary results of this project are described in detail in Zasada *et al.* (2003b) in the proceedings of the joint meeting of the 4th Annual FIA Symposium and 2002 Southern Mensurationists conference. Lowe *et al.* (2003) describe the technical details associated with the GIS analysis of spatial distribution of stream and road buffers and assumptions behind the spatial distribution of the FIA and industrial ground measurement data used in this and the long-term sustainability analysis.

Landscape level inventory visualization: Initial efforts towards effective presentations of the results of inventory projections in the long-term sustainability analysis included research in computer-generated landscape images for inventory

Figure 1.—*Information dissemination on the spatially explicit analysis: A) Phase 1 static presentation; B) Phase 2 dynamic presentation; C) Dynamic map display for various inventory projections; D) Phase 2 of landscape visualization; E) Example of a Phase 2 computer generated image of forecasted inventory.*



visualization. It was a seed project intended to attract explicit funding for such research, results of which could be used in public relations and various analysis of aesthetical, recreational, and visual implications of different forest management practices. Examples of the results of this research have been published on the Web at http://www.growthandyield.com/landscape/ in the form of two tables (e.g., fig. 1D) of image thumbnails. These Web pages include short descriptions of an initial and an advanced phase of the project with a number of examples illustrating how the quality of the images changed at different phases of the project. The quality of the final phase images (e.g., fig. 1E) is by far not the best possible, but the project was suspended due to lack of funding.

Biomass and carbon analysis: We have derived in this study an estimate of carbon and biomass pools in Georgia forests based on the USDA Forest Service Forest Inventory Analysis data of the 1997 survey. The results include tables with estimates and a map of the biomass density and pools at a subcounty level of resolution, which is based on spatially explicit simulations of the potential cover type polygons implied by the FIA data with approximate plot locations. Our results include estimates of the biomass pools in the belowground roots, aboveground tree, and foliage. This study estimates biomass density and pools at a tree level using diameters and heights. The results are then propagated to the plot level using the USDA Forest Service tree expansion factors, and then transformed to plot-dependent polygons using the plot expansion factors. The plot-dependent polygons were spatially simulated using a simplified assumption of homogeneity of conditions surrounding each plot to the extent of the area defined by this plot's expansion factor. Despite the simplified assumption, the derived map provides an excellent visual representation of the distribution of forest biomass densities and pools in the State with distinctive patterns observed in various areas of urban development, federally owned forests, primary commercial forestland, and other land use areas. Coniferous forests with the highest total biomass density are located mostly in three regions: northern Georgia (Appalachian Highlands), the southern part of the Piedmont, and the eastern part of the Coastal Plain. Deciduous and mixed forests with the highest biomass density are concentrated first of all in the northern part of the State—especially in the Blue Ridge physiographic province, and in the western part of the East-gulf Coastal Plain. Counties with the highest biomass density were located primarily in the northern part of the State while counties with the lowest density tended to be on the coast. A journal manuscript describing this study is under review.

Hypermap Web Pages: To facilitate online Internet access to the inventory estimates for various types of geographical divisions in Georgia, we have developed several types of hypermap systems (fig. 2), which are interactive, user-friendly, and rich in abundant information on Georgia forest inventories. The first generation of the hypermaps is available at http://www.growthandyield.com/LiveMaps/clickmaps.htm and requires the user to click on a selection of a display on a first-level selection navigation table. Changing the display of the maps from static to dynamic with different maps displayed as the user moves the mouse over different navigation selections was a major improvement, which can be accessed at http://www.growthandyield.com/LiveMaps/movemaps.htm (fig. 2A). In both these map systems, once the first-level selection is executed, the maps dynamically display a small table with summary statistics for different subunits of the State (fig. 2B). The newest system of the hypermaps is the most comprehensive and dynamic. Mouse-over movement over the first level navigation table dynamically displays tables with various maps that are available for the different divisions of Georgia and can be accessed at http://www.growthandyield.com/movemaps/ (fig. 2C). When one of the map-tables is selected, it becomes a navigation table displaying enlarged map images dynamically as it is browsed over (fig. 2D). When one of the maps is selected, it becomes a navigation table/map displaying dynamically summary graphs for each subunit of the State (fig. 2E). Finally, when the subunit of the State is selected in any of the systems, the user is taken to a page with extensive inventory summary tables (fig. 2F).

Other associated projects and solutions: Among the most important projects associated with the FSA is the geostatistical analysis study described in these proceedings. The objective of this project was to evaluate the applicability of the Landsat TM imagery for analyzing the textural information of pine forest images. Other associated projects have mostly been the results of seeking the best solutions in work organization and the most productive work environment. The bibliography

Figure 2.—*Hypermap Web pages with inventory estimates for different divisions of Georgia: A) Navigation table with dynamic maps; B) Navigation hypermap with a dynamic table; C) Navigation table with dynamic would-be navigation map-table; D) Navigation map-table with dynamic hypermap; E) Navigation hypermap with dynamic graphs.*

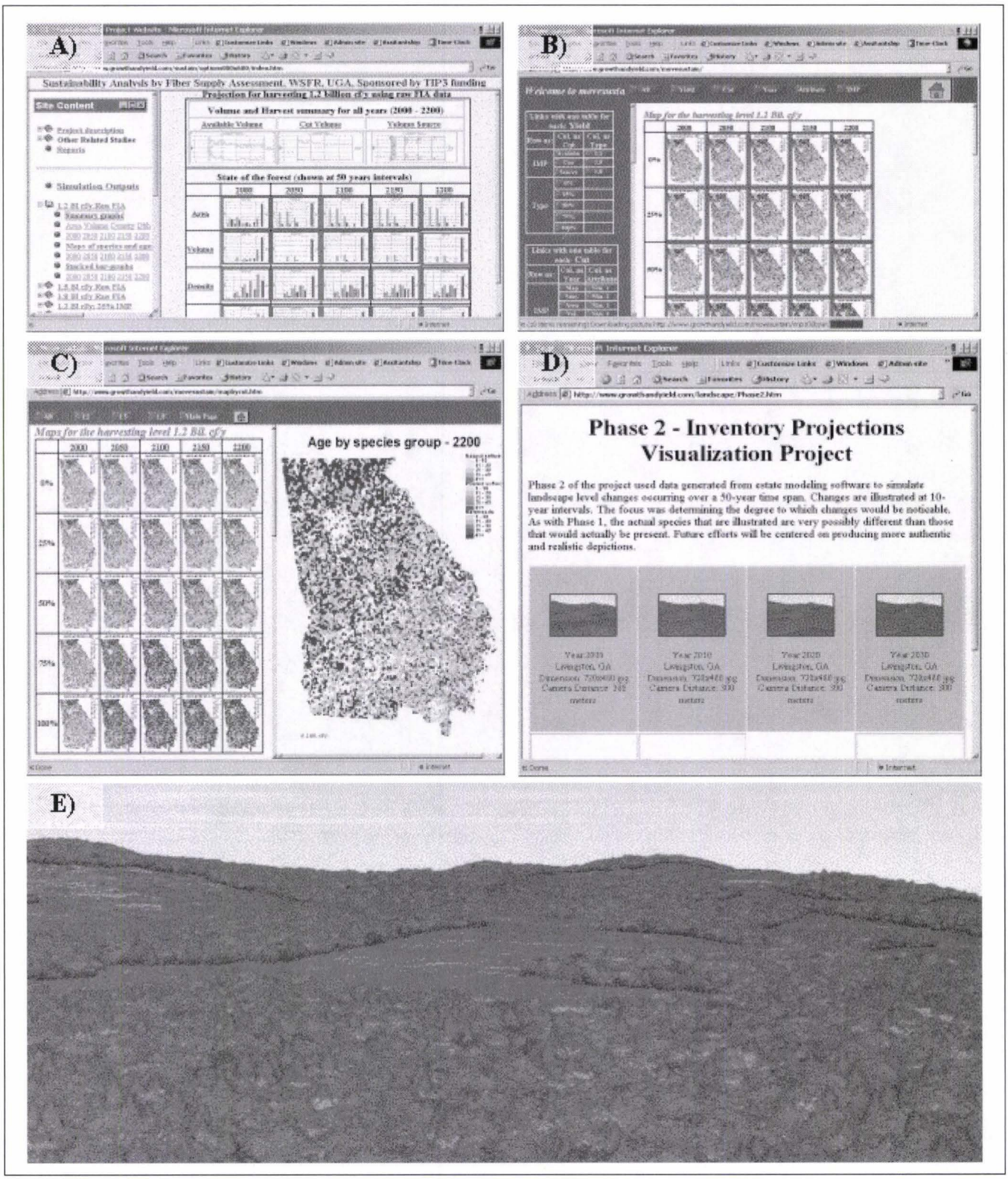

A)
Sustainability Analysis by Fiber Supply Assessment, WSFR, UGA, Sponsored by TIP3 funding
Site Content
Volume and Harvest summary for all years (2000 - 2200)
State of the forest (shown at 50 years intervals)
B)
C)
Maps for the harvesting level 1.2 Bill. cf/y
Age by species group - 2200
D)
Phase 2 - Inventory Projections Visualization Project
Phase 2 of the project used data generated from estate modeling software to simulate landscape level changes occurring over a 50-year time span. Changes are illustrated at 10-year intervals. The focus was determining the degree to which changes would be noticeable. As with Phase 1, the actual species that are illustrated are very possibly different than those that would actually be present. Future efforts will be centered on producing more authentic and realistic depictions.
E)

Figure 3.—*Other studies and administrative/management Web sites associated with the Fiber Supply Assessment program: A) Online literature database; B) Fiber supply assessment online publication retrieval system; C) Comparison of the first panel FIA annual inventory estimates to the last periodic inventory estimates; D) Comparison of the first and second panel FIA annual inventory estimates to the last periodic inventory estimates.*

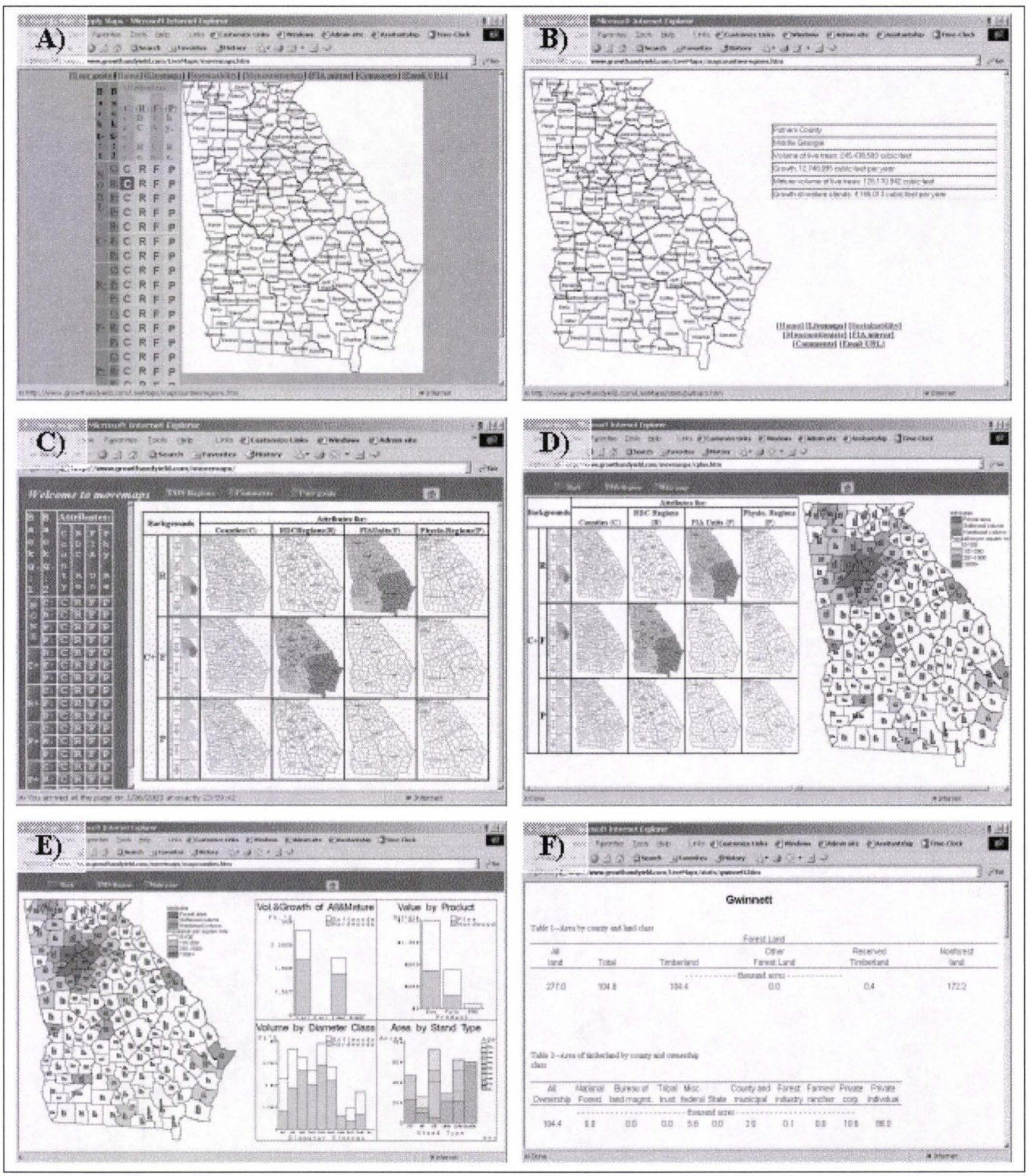
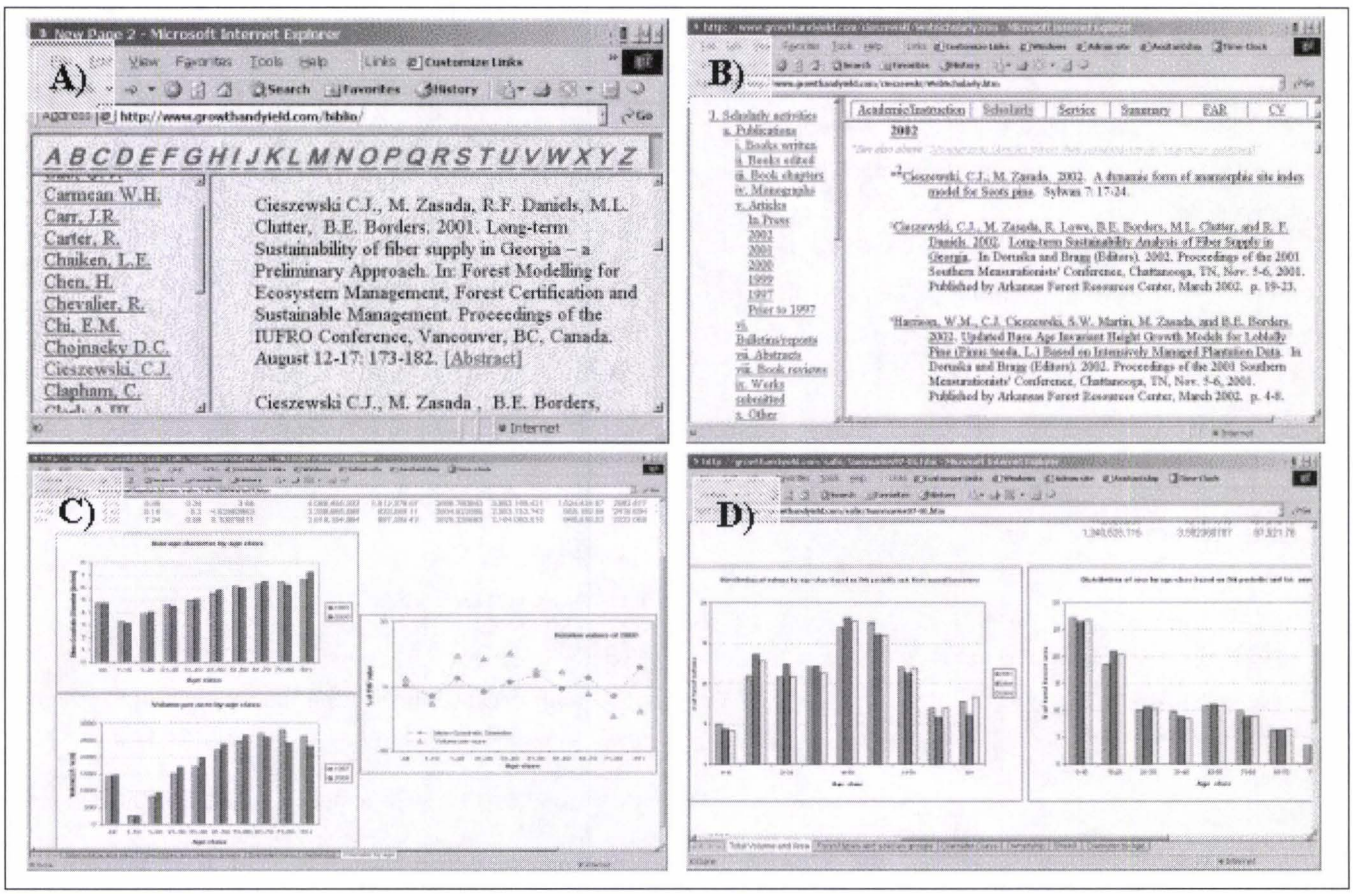


Figure 4.—*Optimized hardware/software solutions: A) Cost of a computing cluster built with inexpensive generic hardware. The lines on the graph mark expandability of the hard disk storage from 1 to 5 Hard Disks per CPU. Note that the 5th HD is priced as an external FireWire disk that is portable and can be used with laptops. B) Mirror of the USDA Forest Service database site, having the same appearance as the original site. C) As in B) but with different skin.*


A)
B)
C)
D)
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Carmean W.H.
Carr, J.R.
Carter, R.
Chniken, L.E.
Chen, H.
Chevalier, R.
Chi, E.M.
Chojnacky D.C.
Cieszewski, C.J.
Clapham, C.
Cieszewski C.J., M. Zasada, R.F. Daniels, M.L. Clutter, B.E. Borders. 2001. Long-term Sustainability of fiber supply in Georgia – a Preliminary Approach. In: Forest Modelling for Ecosystem Management, Forest Certification and Sustainable Management. Proceedings of the IUFRO Conference, Vancouver, BC, Canada. August 12-17: 173-182. [Abstract]
Cieszewski C.J., M. Zasada , B.E. Borders,

project was initiated to streamline literature reviews and database management with online access at http://www.growthandyield.com/biblio/ (fig. 3A). The system contains full texts of all the articles researched in the program (available only for internal use due to copyright laws). All publications produced in the FSA can be retrieved online using a self-maintained user-database at http://www.growthandyield.com/cieszewski/WebScholarly.htm (fig. 3B), or through the information dissemination page at http://www.growthandyield.com/chris/. Finally, the FSA timely comparisons of recent FIA inventory estimates are available online at http://www.growthandyield.com/safis/ in the form of Excel-generated HTML spreadsheets with various associated graphs (e.g., fig. 3C). The spreadsheets contain more extensive comparisons between inventories than those commonly published by the USDA Forest Service or the Georgia Forestry Commission. The main analysis for the comparisons between the inventory estimates is conducted in SAS and then published on the Web using Excel only for presentation of the results.

FSA Innovative Hardware and Software Solutions

Hardware solutions: Analysis of satellite imagery in conjunction with various inventory and GIS data requires massive computing and storage capabilities (needs at the time of writing this piece exceeded 5 Terabytes and were expected to double before June 2003). The most productive and cost efficient solution was based on building a high-speed computing cluster (fig.

4A). We built a 1-Gb project-specific sub-networked computing cluster designated for extensive data processing and forest inventory simulations. The cluster enables the FSA program to perform many new strenuous computational tasks such as long-term, spatially explicit simulations of forest resource availability in Georgia or complex GIS analyses. The project allows for a computer simulation series that previously required months to complete to now be done in less than a week per each subdivision. Moreover, the high-speed sub-network provides opportunities for high-efficiency data storage because transfer rates between different CPUs are comparable to the access rates on native drives. Even with an administrative overhead, the effective time of copying data between different CPUs is much faster than going through the school network. The project-specific sub-network benefits the school at large by freeing up a large amount of resources since all traffic commonly generated by projects is removed from the school lines.

Software "programming optimization": Because many of the FSA studies use the FIA data and its Web pages, it is critical that the relevant online access be in good working order and available at all times. For this reason, we make mirrors of various sites and copy entire structures of their directories onto our computers so that we can use them for offline browsing even when they, or the network, are down. At the same time some Web sites cannot be used offline because of the way they are programmed, and we need then to either reprogram the sites or to refrain from making their mirrors. Such decisions depend on how reliable is the original provider of the subject site and on how much work is required for the reprogramming. The USDA Forest Service Database Retrieval System is a good example of a site that needs reprogramming for offline browsing. First, the provider was frequently inaccessible and even when accessed was painfully slow. The problem is that the program is too advanced. It computes tables in real time generating many temporary files, which quickly reach a critical number—crashing the server. We optimized the program by replacing the dynamic table generating programs with preprocessed static HTML tables. The improved mirror of this database can be accessed at either of the two locations: http://www.growthandyield.com/fiamirror/SE/www.srsfia.usfs.msstate.edu/fia.htm (fig. 4B) and/or http://www.growthandyield.com/fiamirror/ (fig. 4C).

Summary and Conclusions

The FSA has developed many diverse research studies intended to satisfy current funding requirements or to attract different funding sources. Future directions of research will depend chiefly on funding opportunities. In hardware solutions we will follow the computing cluster strategy as the cheapest solution to large computing and storage needs, which can be built even with moderate equipment.

Literature Cited

Bailey, Robert L.; Borders, B.E.; Ware, K.D.; Jones, E.P. 1985. A compatible model relating slash pine plantation survival to density, age, site index, and type and intensity of thinning. Forest Science. 31: 180–189.

Bailey, Robert L.; Clutter, J.L. 1974. Base-age invariant polymorphic site curves. Forest Science. 20: 155–159.

Cieszewski, Chris J. 2001. Three methods of deriving advanced dynamic site equations demonstrated on Inland Douglas-fir site curves. Canadian Journal of Forest Research. 31(1): 165–173.

Cieszewski, Chris J. 2002. Comparing fixed- and variable-base-age polymorphic site equations having single versus multiple asymptotes. Forest Science. 48: 7–23.

Cieszewski, Chris J. [In Press]. Development of a well-behaved dynamic-site-equation based on a modified Hossfeld IV function Y3=axm/(c+xm-1), simplified mixed model, and scant subalpine fir data. Forest Science.

Cieszewski, Chris J.; Bailey, R.L. 2000. Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes. Forest Science. 46: 116–126.

Cieszewski, Chris J.; Bella, I.E.; Walker, D. 1999. Implementation of a base age invariant height model for lodgepole pine in company timber supply analysis. Forestry Chronicle. 75: 1–3.

Cieszewski, Chris J.; Nigh, G.D. 2002. A dynamic equation for a published Sitka spruce site-dependent height-age model. Forestry Chronicle. 78: 690–694.

Cieszewski, Chris J.; Turner, D.P.; Phillips, D.L. 1996. Statistical analysis of error propagation in national level carbon budgets. In: Mowrer, H.T.; Czaplewski, R.L.; Hamre, R.H., eds. Spatial accuracy assessment in natural resources and environmental sciences: 2d International symposium; 1996 May 21-23; Fort Collins, CO: 649–658.

Cieszewski, Chris J.; Zasada, M. 2002. A dynamic form of anamorphic site index model for Scots pine. Sylwan. 7: 17–24.

Cieszewski, Chris J.; Zasada, M.; Daniels, R.F.; Clutter, M.L.; Borders, B.E. 2001. Long-term sustainability analysis of wood supply in Georgia—preliminary approach. In: LeMay, ed. Proceedings, IUFRO conference: Forest modeling for ecosystem management, forest certification, and sustainable management; 2001 August 12–18; Vancouver, Canada: 173–182.

Cieszewski, Chris J.; Zasada, M.; Lowe, R.C.; Borders, B.E.; Clutter, M.L.; Daniels, R.F. 2002. Long-term sustainability analysis of fiber supply in Georgia. In: Doruska; Bragg, eds. Proceedings of the 2001 Southern Mensurationists' conference; 2001 November 5–6; Chattanooga, TN. Arkansas Forest Resources Center: 19–23.

Clutter, Jerome L.; Fortson, J.C.; Pienaar, L.V.; *et al.* 1983. Timber management. New York: John Wiley and Sons, Inc. 333 p.

Coile, Theodore S.; Schumacher, F.X. 1964. Soil-site relations, stand structure and yields of slash and loblolly pine plantations in the Southern U.S.

Lowe, Tripp; Cieszewski, C.J.; Zasada, M.; Zawadzki, J. 2004. Distributing FIA inventory information onto segmented Landsat Thematic Mapper images stratified with industrial ground data. In: Proceedings of the 4th annual Forest Inventory and Analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC–252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Northway, S.M. 1985. Fitting site-index equations and other self-referencing functions. Forest Science. 31: 233–235.

Pienaar, Leon V.; Shiver, B.D. 1986. Basal area prediction and projection equations for pine plantations. Forest Science. 32: 626–633.

Zasada, Michal; Cieszewski, C.J.; Lowe, R.C.; *et al.* 2002. Selected technical aspects of a statewide forest resource-modeling project in Georgia. In: Doruska; Bragg, eds. Proceedings of the 2001 Southern Mensurationists' conference; 2001 November 5–6; Chattanooga, TN. Arkansas Forest Resources Center: 24–28.

Zasada, Michal; Cieszewski, C.J.; Borders, B.E. 2003a. Impact of intensive management practices on long-term simulations of forest resources in Georgia. In: Proceedings of the 2002 SAF national convention. Bethesda, MD: Society of American Foresters. [in press].

Zasada, Michal; Cieszewski, C.J.; Lowe, R.C.; Zawadzki, J. 2003b. Using FIA and GIS data to estimate areas and volumes of potential stream management zones and road beautifying buffers. In Proceedings of the 4th Annual Forest Inventory and Analysis; 2002 November 19-21; New Orleans, LA. Gen. Tech. Rep. NC–252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Rapid Classification of Landsat TM Imagery for Phase 1 Stratification Using the Automated NDVI Threshold Supervised Classification (ANTSC) Methodology

William H. Cooke and Dennis M. Jacobs

Abstract.—FIA annual inventories require rapid updating of pixel-based Phase 1 estimates. Scientists at the Southern Research Station are developing an automated methodology that uses a Normalized Difference Vegetation Index (NDVI) for identifying and eliminating problem FIA plots from the analysis. Problem plots are those that have questionable land use/land cover information. Four Landsat TM scenes in Georgia have been classified using this methodology. A cross-validation approach was used to assess accuracy. The results are compared with an alternative methodology: the Iterative Guided Spectral Class Rejection (IGSCR) methodology.

Several FIA units have examined methodologies that test the usefulness of pixel-based estimates for Phase 1 stratification. Among these are k-Nearest Neighbor (k-NN) (Franco-Lopez *et al.* 2000), Iterative Guided Spectral Class Rejection (IGSCR) (Wayman *et al.* 2001) and various model-based approaches (Moisen *et al.* 1998). A new methodology developed by scientists at the USDA Forest Service Southern Research Station seeks to combine simple concepts of satellite image data classification with FIA plot data and automate the process. This new methodology compares FIA plot information with spectral information from an NDVI transform, using an automated approach for choosing Euclidean distances used to generate FIA plot-based classification "signatures." An additional component of this methodology was tested that examines crown modeling quantitatively to assess the usefulness of FIA plots for generating signatures over the portion of the NDVI range (150–185) that is most problematic for distinguishing forest from nonforest pixels. The result of these comparisons is the development of efficient Phase 1 classification techniques that meet FIA remote sensing business requirements.

Operational Efficiencies

The Southern Research Station inventories forests in 13 Southern States and requires approximately 131 TM scenes for complete "wall-to-wall" coverage of all States. Phase 1 stratification procedures need to keep pace with changes in forest conditions in the South and with the pace of inventory reporting cycles that require re-measuring all FIA ground plots every 5 years. The rate of change of southern forests is rapid and subject to environmental, social, and economic forces including:

- Clearcutting
- Urbanization
- Landowner assistance programs
- Population shifts

Any classification methodology adopted for FIA should be operationally efficient for FIA purposes and address the following requirements:

- High automation potential
- Straightforward implementation
- High CPU and storage efficiencies
- High repeatability

To date, the various Phase 1 methodologies that have been proposed and tested have failed to meet one or more of these requirements. For example, the IGSCR methodology requires a great deal of subjective interpretation to establish signatures and the iterative nature of the classification requires a great deal of storage space.

Figure 1 indicates the study area for the ANTSC methodology test project. Figure 2 indicates the subset of the study area used for examining crown modeling approaches aimed at refining the NDVI threshold component of the ANTSC methodology. Comparison of the results of the ANTSC methodology with the IGSCR methodology requires examining both methodologies in more detail.

Figure 1.—*Study area for ANTSC methodology test project.*

Figure 2.—*Subset of the study area used for tests of crown modeling.*

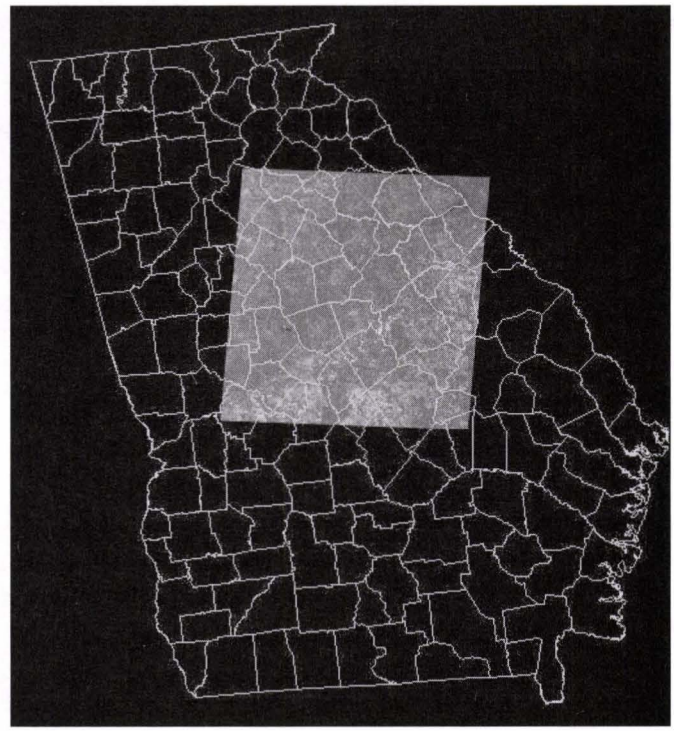
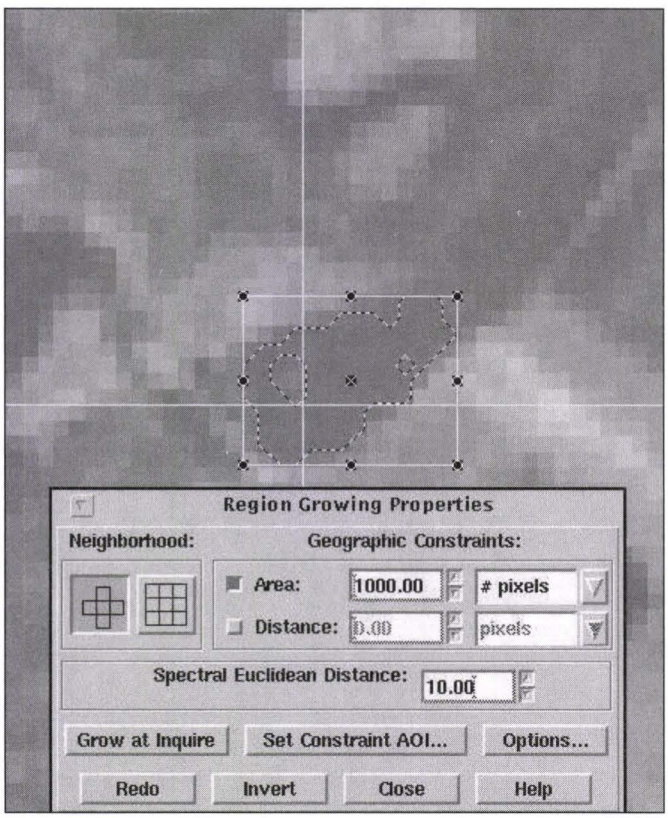


Figure 3.—*Region that was grown using a Euclidean distance of 10.*



IGSCR Methodology

The IGSCR methodology uses FIA plot information for developing statistical signatures. These signatures consist of the mean and variance of the spectral reflectance of the ground conditions in several Landsat TM spectral channels. The analyst views the location of the FIA plot on the image and, at that spot, chooses a pixel (seed) for the signature growing process. Using the pixel collocated at the FIA plot position, the analyst specifies a Euclidean distance in multi-spectral space that captures contiguous pixels to be accepted, if within the Euclidian distance of the same land use condition. Pixels outside the distance are rejected as the same land use condition. The analyst must be able to recognize whether the region included in the signature growing process remains in the land use condition of seed pixel initiation. Figure 3 indicates a region that was grown using a Euclidean distance of 10. The analyst must adjust the Euclidean distance to ensure that the signature does not grow beyond the land use class of initiation, so must frequently zoom in and out of the image to subjectively assess the results of the seed-growing process.

Table 1.—*Georgia IGSCR/ANTSC accuracy assessment comparisons*

TM scene path/row	Forest				Nonforest		
	Overall accuracy	Producers accuracy	Users accuracy	Kappa statistic	Producers accuracy	Users accuracy	Kappa statistic
IGSCR							
17/37	84.79	89.67	88.71	0.6241	73.42	75.32	0.6473
17/38	85.38	92.44	89.08	0.5516	65.28	75.20	0.6649
18/37	84.93	92.17	89.08	0.4855	58.06	66.67	0.5768
18/38	86.71	98.22	83.73	0.5519	66.52	95.51	0.9296
ANTSC							
17/37	90.01	91.03	97.26	0.9884	90.48	73.08	0.7498
18/37	95.28	95.11	99.43	0.9570	96.43	75.00	0.7120
18/38	95.01	95.52	99.58	0.9884	99.32	88.48	0.8182

The IGSCR process is detailed in Wayman (2001). To begin the IGSCR classification process, an unsupervised classification of 100 classes using a convergence threshold of 0.95 and variance set to one standard deviation was performed for each TM image. Collected signatures were then used to extract the class values that result from the classification process, and output those class pixel values to a text file suitable for statistical analyses. The class information was analyzed for purity (95 percent) and classes deemed pure were removed (masked) from the original TM imagery. The remaining image pixels were then separated into 100 classes for the second iteration of class purity testing. At least three iterations were performed for each image.

Table 1 lists the accuracies obtained for each of the four TM scenes that were classified using the IGSCR methodology. The methodology was relatively accurate for the binary classification of the forest and nonforest conditions, but required significant analyst time and effort for choosing Euclidean distances in the signature collection process. The multiple classifications of the imagery required by IGSCR occupied a lot of storage space. These shortcomings of the methodology prompted the development of a hybrid classification approach combining NDVI-based techniques (Hoppus *et al.* 2000) with the Euclidean distance signature development component of the IGSCR methodology.

ANTSC Methodology

The IGSCR subjective signature generation process relies on visual interpretation of forest and nonforest cover types. Familiarity with the landscape and ecosystem processes is a prerequisite for accurate image classification. At present, the signature collection process is time consuming and tedious, and interpreter fatigue is a real problem.

Euclidean Distance Component

Signature collection in support of the IGSCR methodology resulted in the visual interpretation of over 1,200 signatures for four TM images from 1992 and four TM images from 2000. These results suggested that a Euclidean distance of 13 optimized signature growth for forested conditions but rarely caused the signature to grow out of the condition of seed pixel initiation. A Euclidean distance (D) of 21 gave similar results in nonforest conditions.

Euclidean distance, D:

$$D_{ab} = \left[\sum_{i=1}^{n} (a_i - b_i)^2 \right]^{.5}$$

Where a and b are values of pixels being evaluated and n is the total number of satellite layers.

NDVI Component

A large body of literature exists confirming the usefulness of the Normalized Difference Vegetation Index (NDVI) band transformation for extracting information about forest vegetation (Iverson *et al.* 1989, Anderson *et al.* 1993). Results using an NDVI threshold by the Northeast FIA unit confirmed that NDVI was useful for separating forest from nonforest conditions. Figure 4 illustrates how the NDVI values for FIA plots (subplot 1) compare for a single TM scene.

The search for operationally efficient automated classification methodologies led researchers at the Southern Research Station (SRS) to develop an integrated methodology that used an NDVI threshold with automated signature collection to rapidly classify TM images using a Maximum Likelihood-based "Supervised Classification" approach, dubbed the **A**utomated **N**DVI **T**hreshold **S**upervised **C**lassification (ANTSC) method.

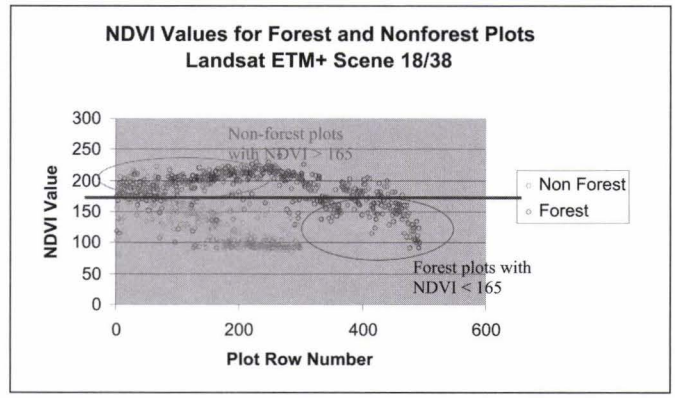
An NDVI threshold of 165 was used to differentiate between forest and nonforest. Each FIA plot's NDVI value was extracted from an NDVI-transformed TM image using a Pixel-to-ASCII extraction program. The NDVI values were compared to the field-derived land use information. Forested plots with NDVI values below 166 and nonforest plots with NDVI values above 165 were considered separate populations of plots that did not represent land cover information contained with the spectral response surface of Landsat TM imagery. Several explanations for the origination of this population of plots may be hypothesized. The following are possible:

- Change based on disturbance
- Land use versus land cover differences (clearcut = forest)
- Pixel/plot mis-registration

It was considered important to the ANTSC process that this NDVI or parity test be conducted to remove these plots from training and accuracy assessment. Certainly, the removal of these plots purifies the training and accuracy assessment pool of plots used in the cross-validation approach. The IGSCR methodology also indirectly purifies the accuracy assessment pool of plots by removing those plots that resulted in poor signatures during the signature generation process. A poor signature was one that did not include a minimum of 9 pixels, or one that grew into a land cover class different from that of the original pixel.

Accuracies for three TM scenes classified using the ANTSC methodology are shown in table 1. A final accuracy assessment test was performed using the accuracy assessment plots from each method (IGSCR, ANSTC) to test the accuracy of the other method. Results showed accuracy differences for the three scenes done by both methods to be less than 5 percent. Differences in operational efficiency between the two methods were obvious. The IGSCR method took 3 to 7 days per scene, while the ANTSC method took less than 1 day. It should be noted that working through the IGSCR methodology enabled the automated specification of Euclidean distances for the ANTSC methodology. It is not known whether the specification of Euclidean distances for forest and nonforest used in this study are stable across a wide variety of ecological conditions or differing image radiometric conditions.

Figure 4.—*NDVI values for FIA plots on a Landsat TM scene at FIA subplot 1.*



Utilizing a hard NDVI threshold of 165 assumes that the NDVI ratio is consistent from image to image and that radiometric differences among images are not reflected in the NDVI transform. To test the concept of using a soft threshold, plots that fell into the range of NDVI values between 145 and 165 were assessed for their correct land use call by using a process of crown modeling. Crown modeling uses the distance and azimuth of each tree tallied on an FIA plot, coupled with regression estimates of crown width derived from Forest Health Monitoring (FHM) data, to calculate the proportion of crown reflectance per FIA subplot. These subplot proportions were compared with the NDVI values at the same location to determine land use/land cover compatibility. A somewhat arbitrary threshold of 16.7 percent crown cover per FIA subplot was chosen as the cutoff between forest and nonforest conditions for the comparisons made in this study.

Crown Modeling

Crown modeling for calculating the average canopy reflectance by subplot follows these steps:

- Develop local regressions that predict crown diameter by species from Forest Health Monitoring data.
- Compute the crown radii for each tree species.
- Use a buffer approach in the GIS software to draw the crowns in their real world locations.
- Intersect the crowns with the subplot circles and calculate proportional reflectance per subplot/plot.

For the subset study area of one TM scene, 28 FIA plots fell within the 145-165 NDVI range. Of these 28 plots, 4 had crown proportion reflectance percentages that were inconsistent with the FIA land use call. The crown models are superimposed on the TM imagery and comparisons shown for 3 of these plots in figures 5, 6, and 7.

Figure 5 shows that for this FIA plot, subplot 4 fell in a forest. The average crown proportion for the four subplots was 19.4 percent. This exceeds the 16.7 percent threshold of canopy reflectance, but the NDVI value (154) for this plot was determined from the pixel that corresponded to subplot 1. Since the calculated average crown reflectance proportion was inconsistent with the NDVI value at subplot 1, the analyst has the option to use the pixel at subplot 1 as a seed for a nonforest signature since the plot was not thrown out on the basis of the NDVI parity test.

Figure 6 shows an FIA plot that is classified as forest in the field, but the calculated average crown reflectance proportion (13.8 percent) is less than the 16.7 percent threshold. The crown models reveal a plot that is in an area that was likely clearcut a few years ago and is reverting to forest. The canopies are small and the crown reflectance proportion calculations are predicated on using FIA tally trees that are 5 inches

Figure 5.—*Crown proportion (19.4 percent) and NDVI value (154) not consistent.*

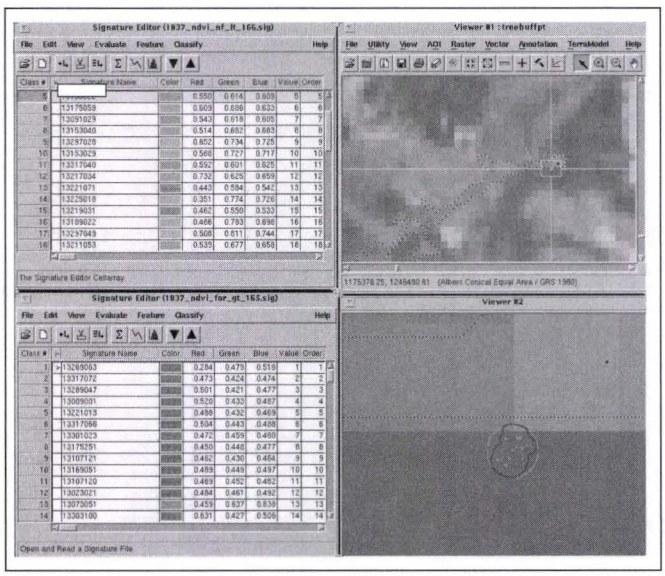


Figure 6.—*Crown proportion (13.8 percent) and NDVI value (174) not consistent.*

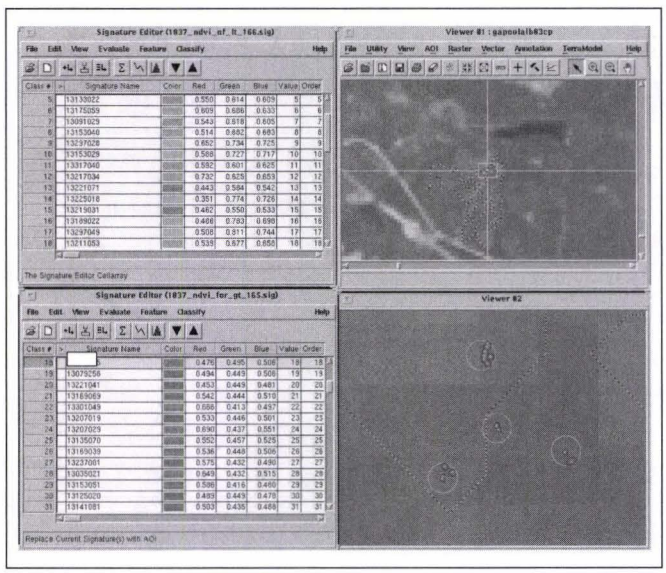
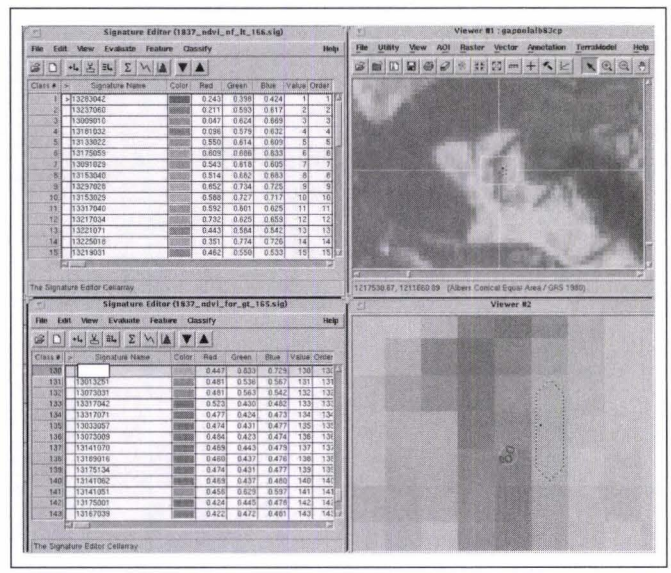


Figure 7.—*Crown proportion (7.9 percent) and NDVI value (170) not consistent.*



d.b.h. or greater. The NDVI value at subplot 1 (174), indicates a forest condition that is consistent with the FIA land use call but inconsistent with the crown modeling-based proportion. The analyst should initiate a seed based on the NDVI value that is consistent with the land use call in the field.

Figure 7 shows an FIA plot in a recent clearcut that has a low average crown reflectance proportion (7.9 percent) but a relatively high NDVI value at subplot 1. It is obvious that subplot 1 falls in a forest edge while the other 3 subplots fall in the clearcut (nonforest). The crown modeling procedure points out a classic land use/land cover conflict. If the analyst places the seed for this forested plot at subplot 1, the signature will reflect the nonforest condition. If the analyst places the seed for this forested plot at subplot 4, the signature will reflect the forested condition. In this case, the crown proportion calculations raised a red flag that leads the analyst to a closer look at the land use/land cover issue.

Conclusions

Classification accuracies for the ANTSC and the IGSCR methodologies were similar. The ANSTC classification methodology is less subjective and requires no analyst input, making it easy to implement by analysts with minimum remote sensing expertise. Results of the crown modeling experiments indicate that the NDVI threshold of 165 is a good choice but some land use/canopy reflectance inconsistencies exist with the 145-165 NDVI range. The number of inconsistencies was small (<14 percent of the total FIA plots). The additional time spent assessing the problem plots within the 145-165 NDVI range is likely worth the improvement in precision, although a small amount of automation potential may be sacrificed.

It is not known whether the Euclidean distance measures used in the ANTSC methodology will work as well in other States or in different ecological conditions. It is possible that some preliminary work will be required to determine the optimum Euclidean distances for forest and nonforest signatures when ecological conditions are significantly different.

Literature Cited

Anderson, G.L.; Hanson, J.D.; Haas, R.H. 1993. Evaluating Landsat thematic mapper derived vegetation indices for estimating above ground biomass on semiarid rangelands. Remote Sensing of the Environment. 45: 165–175.

Franco-Lopez; H.; Ek, A.R.; Bauer, M.E. 2000. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of the Environment. 77(2): 251–274.

Hoppus, M.; Riemann, R.; Lister, A. 2000. Remote sensing strategies for forest inventory and analysis utilizing the FIA plot database. In: 8th Forest Service biennial remote sensing applications conference; ASPRS, Albuquerque, NM, April 2000.

Iverson, L.R.; Cook, E.A.; Graham, R.L. 1989. A technique for extrapolating and validating forest cover across large regions calibrating AVHRR data with TM data. International Journal of Remote Sensing. 10(11): 1805–1812.

Moisen, G.G.; Edwards, T.C.; Van Hooser, D. 1998. Merging regional forest inventory data with satellite-based information through nonlinear regression methods.

Wayman, J.P.; Wynne, R.H.; Scrivani, J.A.; Reams, G.A. 2001. Landsat TM-based forest area estimation using iterative guided spectral class rejection. Photogrammetric Engineering and Remote Sensing. 67(10): 1155–1166.

Quantifying Forest Ground Flora Biomass Using Close-range Remote Sensing

Paul F. Doruska; Robert C. Weih, Jr.; Matthew D. Lane[1]; and Don C. Bragg[2]

Abstract.—Close-range remote sensing was used to estimate biomass of forest ground flora in Arkansas. Digital images of a series of 1-m^2 plots were taken using Kodak DCS760 and Kodak DCS420CIR digital cameras. ESRI ArcGIS™ and ERDAS Imagine® software was used to calculate the Normalized Difference Vegetation Index (NDVI) and the Average Visible Reflectance (AVR) index for each plot. Regressions, developed to estimate green and dry biomass from the NDVI and/or AVR values, explained 30-40 percent of the variation. A vegetation mask and/or different independent variables are needed to improve the regression models.

Many forest research projects estimate forest ground flora biomass via the labor-intensive technique of clipping, drying, and weighing vegetation samples (Brower *et al.* 1990). When combined with species identification, such work is used to report various diversity measures (Elzinga *et al.* 1998, Foti and Devall 1994, Magurran 1988) used in ecosystem studies and reporting. Such information is also used when assessing wildlife habitat (National Wildlife Research Center 2000).

Satellite imagery combined with computer algorithms has been used to estimate forest biomass (Ahern *et al.* 1991, Baret *et al.* 1989), but this imagery cannot be used to estimate forest ground flora biomass because of canopy blockage and scale. Our pilot project sought to determine whether techniques used to estimate forest biomass from satellite imagery can be used to estimate forest ground flora biomass using close-range, remotely sensed imagery. Photoplots have been used in ecological research for change detection (Schwegman 1986, Windas 1986). This project combines the use of photoplots with the techniques of satellite imagery to estimate forest ground flora aboveground biomass.

Equipment

Two digital cameras were used in conjunction on this project: A Kodak DCS760 camera with a Nikon F5 body was used to take color digital images at a 6 million pixel (3038 x 2028) resolution; a Kodak DCS420CIR camera with a Nikon F90 body camera operating at a 1.5 million pixel (1524 x 1020) resolution was used to take the color infrared images. Twenty mm auto-focus lenses were used on both cameras, and an Omega Optical band pass filter (500-900 nm) was used in conjunction with the DCS420CIR camera to block blue light.

An aluminum stand was constructed to frame the 1-m^2 plot and mount the cameras. (The actual frame size was 0.966 m^2 but will be referred to as 1 m^2 in this manuscript.) The cell size for the imagery was 0.1015 cm. Black and white bands were painted onto the frame to calibrate images from 0 to 255 to take into account variations in illuminations. Cross hairs (or tick marks) were drawn onto the frame to develop a local coordinate system for image comparisons. ESRI ArcGIS™ 8.x and ERDAS Imagine® 8.5 software was used to process the imagery and calculate the vegetative indices used herein.

Methods

Collecting Images and Vegetation

A series of 1-m^2 plots were randomly established on the University Forest at the University of Arkansas-Monticello for the initial analysis. Once a plot was located, the aluminum camera stand was set up and vegetation overlapping or extending beyond the border of the frame was removed to ensure only vegetation within the plot would appear in the images. Each camera was then mounted on the frame separately and raised to the appropriate level. Three pictures were taken per camera to be sure at least one usable image was captured. After the images were taken, the vegetation on the plots was clipped at

[1] Assistant Professor, Spatial Information Systems Program Director, and Research Specialist, respectively, Arkansas Forest Resources Center, University of Arkansas School of Forest Resources, Monticello, AR. E-mail: doruska@uamont.edu.
[2] Research Forester, U.S. Department of Agriculture, Forest Service, Southern Research Station, Monticello, AR. E-mail:dbragg@fs.fed.us.

ground level, sorted by species, and placed into labeled plastic bags and sealed for laboratory analysis.

Mass Determination

The green mass of the contents of each bag was determined immediately upon return to the laboratory. The contents of each plastic bag were then transferred to labeled paper sacks and placed in a drying oven at 60°C for 3 to 4 days. After drying, the dry mass of the contents of each bag was measured.

Image Registration and Standardization

The camera stand used in this study had seven tick marks on its frame. These tick marks were measured to within 0.025 cm and placed in a shapefile to represent a local coordinate system for the camera stand. Each collected image was then registered to that coordinate system within ArcGIS™ ArcMap™ using the georeferencing extension. The referencing was done by lining up the measured tick marks with the marks seen on the image; this ensured that all images would line up exactly with each other and could be compared.

Because the amount of solar energy incident on the plots can change, the camera stand also had black and white painted regions on it so the illumination of images could be standardized. To ensure that the digital values represented the same color from one image to the next, a GER2600 spectroradiometer determined the reflectance of the painted regions for four bands (Near Infrared[NIR], Red[R], Green[G], and Blue[B]) and represented the extremes of the range of colors present in any image for any band. A simple linear regression was created per band per image to convert the range of values present within a given band/image combination to the range defined via the spectral radiometer. The regressions were used to calibrate each image.

Images were then subsetted by creating areas of interest (AOI's) manually in Imagine®. The AOI's contained only that portion of each image that was inside the borders of the camera stand and were used for all subsequent analyses.

When applying the regression models to the areas of interest for each band/image combination, any values in the output grid less than 0 were reset to equal 0 (negative values can disrupt calculation of certain vegetation indices). The final output grid was a six-band image consisting of standardized NIR, red and green bands from the color infrared image, and the red, green, and blue bands from the color image.

Vegetation Indices and Regression Modeling

The CIR camera images were used to calculate the Normalized Difference Vegetation Index (NDVI). The Average Visible Reflectance (AVR) index was calculated using the color camera images for each pixel in each image.

$$NDVI = (NIR-R)/(NIR+R) \quad (1)$$

$$AVR = (G+R+B)/3 \quad (2)$$

The output image from this step was a two-band grid (NDVI and AVR) with a cell size of 0.1015 cm by 0.1015 cm.

Once the NDVI and AVR values for the images were calculated, they were summed and averaged for use as potential independent variables in regression equations to predict green or dry biomass.

Results and Discussion

Thirteen of the plots have been completely processed so far. Green mass of the forest floor aboveground vegetation ranged from 30 g to about 415 g and dry mass ranged from 14 g to about 200 g. Figures 1 and 2 show the relation between green and dry biomass, respectively, versus the sum of the NDVI values of the plots. A slight curvilinear pattern is apparent.

Several regression model forms were examined to fit a curve to the data appearing in figures 1 and 2. For predicting or estimating mass (green and dry, respectively) in grams, the fol-

Figure 1.—*Relationship between green mass (g) and NDVI values.*

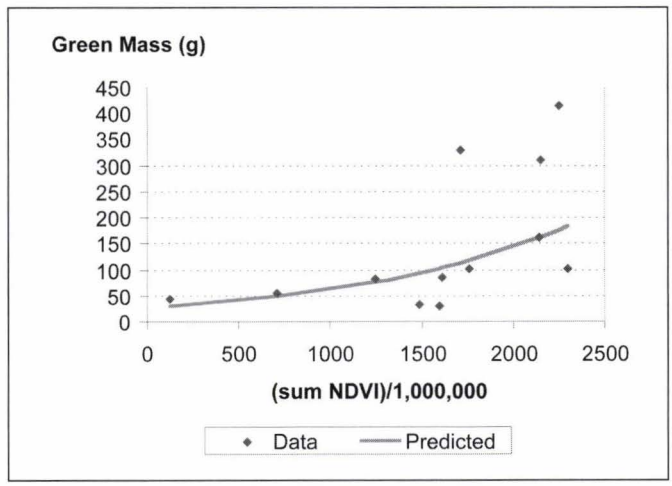
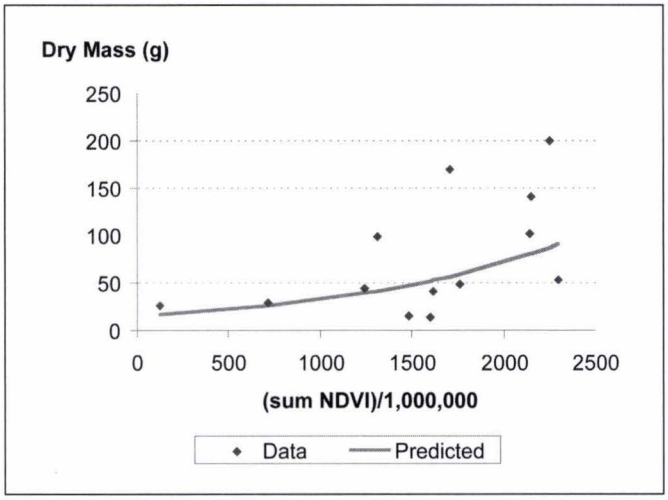


Figure 2.—*Relationship between dry mass (g) and NDVI values.*



lowing model form was most successful across the collection of images (i's):

$$Ln(mass_i) = \beta_0 + \beta_1(\text{sum NDVI}_i) + \varepsilon_i \quad (3)$$

Fit statistics for equation (3) appear in table 1.

NDVI was found to be a better independent variable than AVR. Although parameter estimates were significant, the R^2's of these initial models were fairly low. A few plots greatly impacted model performance. Visual inspection of the images of these plots indicated that a fair amount of vegetative overlap was present, which prevents the cameras from seeing the true quantity of vegetation. Vegetation overlap is definitely of concern to the researchers. If the problem persists as the data set grows, a method to account for vegetative overlap needs to be developed and included.

Other model forms/variables will be considered as the data set continues to grow. In this initial analysis, calculated vegetative indices (NDVI and AVR), not individual color bands, were used when building regression models. Use of individual color band values, especially red, may improve model performance. A vegetation mask was not used in this initial analysis, but will be used in any future analyses. We hope a vegetation mask will further distinguish pixels that contain live vegetation from pixels that contain just the forest floor.

These preliminary results suggest the potential of handheld color and color infrared cameras for quantitative forest floor vegetation sampling by means other than clipping and weighing. This project, as it unfolds, should serve as a good first step in that direction.

Acknowledgment

The authors would like to thank the National Science Foundation (BDEI/0131801) and the Arkansas Forest Resources Center for financial support of this project. This manuscript is approved for publication by the Director, Arkansas Agricultural Experiment Station.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Literature Cited

Ahern, F.J.; Erdle, T.; McLean, D.A.; Kneppeck, I.D. 1991. A quantitative relationship between forest growth rates and thematic mapper reflectance measurements. International Journal of Remote Sensing. 13(4): 699–714.

Table 1.—*Fit statistics for equation 3*

	Parameter	Estimate	Std. error	p-value	R2
Green mass	β_0	3.2934	0.5833	0.0002	37.40%
	β_1	8.3450×10^{-10}	3.4143×10^{-10}	.0346	
Dry mass	β_0	2.6745	0.6078	.0013	33.40%
	β_1	7.9650×10^{-10}	3.5574×10^{-10}	.0491	

Baret, F.; Guyot, G.; Major, D.J. 1989. TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In: 12th Canadian symposium on remote sensing and IGARSS'90; 1989 July 10–14; Vancouver, Canada: 1355–1358.

Brower, J.E.; Zar, J.H.; von Ende, C.N. 1990. Field and laboratory methods for general ecology. Dubuque, IA: William C. Brown Publishers. 237 p.

Elzinga, C.L.; Salzer, D.W.; Willoughby, J.W. 1998. Measuring and monitoring plant populations. Washington, DC: Bureau of Land Management, U.S. Government Printing Office. 477 p.

Foti, T.; Devall, M.S. 1994. Herbaceous plant biodiversity of stands in the Ouachita and Ozark National Forests. In: Proceedings of the symposium on ecosystem management research in the Ouachita Mountains: pretreatment conditions and preliminary findings. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Research Station: 50–60.

Magurran, A.E. 1988. Ecological diversity and its measurement. Princeton, NJ: Princeton University Press. 179 p.

National Wildlife Research Center. 2000. *web;* www.nwrc.gov/wdb/pub/hsi/hsiindex.htm. Accessed 2002 December 1.

Schwegman, J. 1986. Two types of plots for monitoring individual herbaceous plants over time. Natural Areas Journal. 6: 64–66.

Windas, J.L. 1986. Photo-quadrat and compass mapping tool. Natural Areas Journal. 6: 66–67.

Measuring Forest Area Loss Over Time Using FIA Plots and Satellite Imagery

Michael L. Hoppus and Andrew J. Lister[1]

Abstract.—How accurately can FIA plots, scattered at 1 per 6,000 acres, identify often rare forest land loss, estimated at less than 1 percent per year in the Northeast? Here we explore this question mathematically, empirically, and by comparing FIA plot estimates of forest change with satellite image based maps of forest loss. The mathematical probability of exactly estimating a 5-percent loss within a 600,000-acre forest, where 5 percent has actually been converted, is 18 percent. A GIS experiment in Connecticut, using 452 FIA plots and a satellite-derived forest cover map, where 5 percent of the total forest area was "lost" by 7.5-acre units, indicates that the sample estimates a 5-percent loss 35 percent of the time with a range of estimated loss of 3 to 8 percent. Satellite image classification can probably estimate the amount of forests lost to urbanization more accurately, especially over small areas, while providing a more useful map of forest loss.

Forest Inventory and Analysis (FIA), a program of the USDA Forest Service, is responsible for the nationwide forest inventory and monitoring of the United States. Congress mandates, through the Forest and Rangeland Renewable Resources Planning Act of 1974 and the McSweeney-McNary Forest Research Act of 1928, that FIA continuously determine the extent, condition, and volume of timber, as well as the growth and depletion on the Nation's forest land. FIA inventories must meet specified sampling errors: a 3-percent error per 1 million acres of timberland is the maximum allowable for area. Timberland is a category of forest land that is producing or capable of producing 20 cubic feet of industrial wood per acre per year (Hansen *et al.* 1992). Timberland area is usually 80 percent or more of the total forest land of the States in the Northeast.

Users of FIA data in the Northeast increasingly are interested in how much forest land is being converted to other land uses, such as residential housing developments. A low rate of forest land loss per year can still amount to an ecologically significant area. An annual 0.3 percent loss of forest land in the Northeast would equal about 436 square miles, or 2,180 square miles of forest land in 5 years (1.5 percent). The question is whether the current FIA inventory program can identify small rates of forest loss when it occurs somewhat randomly over large areas.

Another source of survey information that the FIA program is evaluating is Landsat 7 satellite imagery. Imagery from two dates can provide a forest loss map. The maps have no sampling error but do have classification or mapping errors. The question is whether the accuracy of such maps is sufficient to determine how much forest is being lost.

Objectives

One objective of this study was to see if it is reasonable to reject the null hypothesis that the density of FIA plots (1 plot per 6,000 acres) does not permit accurate estimates of forest land loss, if the area of loss is only 1 to 5 percent of the total forest area. This hypothesis was evaluated both in terms of mathematical probability and by an empirical GIS evaluation of how actual FIA plots located on a real forest-cover map identify forest loss that was artificially induced at random with a computer.

Another objective was to evaluate the ability of Landsat satellite imagery to detect and map forest loss. Several change-detection methods were evaluated and the resulting change maps compared with the corresponding estimates of forest change made by the FIA ground plot survey. The ultimate goal of the investigators was to evaluate the ability of the FIA survey design to estimate the rare occurrence of forest loss compared to total forest cover in the Northeast, and to determine

[1] Research Forester and Forester, respectively, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA 19073. Phone: 610–557–4039; fax: 610–557–4250; e-mail: mhoppus@fs.fed.us.

how well satellite-derived change-detection maps could provide additional forest loss information.

Methods

Determining the Binomial Probabilities of Forest Loss Given the General FIA Sampling Design

To determine the likelihood that the FIA survey design identifies rare patches of forest loss, a binomial probability function was applied to 100 FIA ground plots within a 600,000-acre forest in which 5 percent of the forest area has been cleared at random in circular patches of 5 acres. For simplicity, we assume that each FIA ground plot consists of a 1-acre circle. (An actual plot consists of four 0.04-acre circular subplots distributed evenly within a 1-acre circle.) The 600,000-acre forest consists of 100 hexagons covering 6,000 acres each. Each hexagon contains one FIA ground plot.

The probability that a plot will hit a cleared patch exactly r times, P(r), is given by the binomial probability function:

$$P(r) = \left(\frac{n!}{r!(n-r)!} \right) p^r (1-p)^{n-r}$$

where:

- p = the probability of a plot hitting a forest loss patch on a single trial
- n = the number trials
- r = the total number of times that plots hit any forest loss patch

A hit occurs when a plot center falls within any portion of the 5-acre circular patches of cleared forest. The numbers represented by p and n were computed by throwing 6,000 5-acre patches of "forest loss" at random into the 600,000-acre forest that contains 100 plots (1 plot per 6,000 acres).

The formula presented here is not precisely correct because it applies to sampling with replacement. It unrealistically allows two forest loss patches to be located on the same piece of ground. However, this formula still accurately approximates the probability that the total number of times a plot hits a forest loss patch equals r, since n, the number of trials, is small compared to the total number of trials possible (Huntsburger and Billingsley 1977). The formula that deals with sampling without replacement is difficult to use with large differences between p and $(1-p)$—hypergeometric probability is generally applied to much smaller problems.

GIS Experiment Using FIA Plot Locations in Connecticut and a Forest Cover Map of the State

A more empirical approach to determining how well the FIA plots estimate forest loss was used to experiment with various levels of loss in a more realistic setting. Using GIS software, we combined the actual FIA plot locations with a forest/nonforest map produced from the U.S. Geological Survey Multi-Resolution Landscape Characterization (MRLC) vegetation cover map for Connecticut derived from a classified 1993 Landsat TM satellite image. Connecticut had about 1,825,700 acres of forest land (59 percent) out of a total land area of 3,117,800 acres (Brooks *et al.* 1993). Plots located in forested areas, according to the MRLC map, totaled 300 out of 452. Five levels of forest loss were applied to the forest area of the map: 1, 2, 3, 4, and 5 percent. Each level of loss was applied at random in multiples of 7.5-acre units. For example, a 1 percent loss in forest land in Connecticut (18,257 acres) required 2,434 unit areas of loss at 7.5 acres each. Random selection of forest loss was repeated 250 times, and the number of forested plots changing to nonforest counted for each percentage level of loss. A histogram of the 250 counts of the number of plots hitting an area of forest loss was plotted for each of the loss levels. In this experiment, a hit occurred when the center of a previously forested plot fell on any portion of an area of forest loss. If the plots can capture low rates of loss, the expected value (mean), or highest frequency in the histogram (mode), should be equal to the percent loss applied to the forested map area.

A Comparison Between Change (Loss) Detection Maps Produced from Landsat TM Satellite Imagery and Estimates of Forest Change Provided by the FIA Survey Reports

Change detection maps derived from two Landsat satellite images, taken 5 to 6 years apart, were evaluated for their ability to estimate forest land area that had changed to other land cover types. If accurate, these maps also could show where this change is taking place and the size class distribution of the change. First, a change map of a large portion (64 percent) of

New York State was evaluated. This map, commissioned by the National Oceanic & Atmospheric Administration (NOAA) was produced by Earth Satellite Corp. (Earthsat). The NOAA Coastal Change Analysis Program (C-CAP) is a national effort to develop and distribute regional land cover and change analysis data for the coastal zone (including the Great Lakes) by using remote sensing technology. C-CAP classifies land cover types into 22 standardized classes that include forested areas, urban areas, and wetlands (fig. 1). For New York, the changes between two dates of satellite imagery, 1995 and 2000, are based on how these 22 classes of land cover are converted from to another over the 5-year interval. There are five classes of forest land: Deciduous, Evergreen, Mixed, Palustrine Forested Wetland, and Estuarine Forested Wetland. In the C-CAP classification scheme, these forest classes can change to: High Intensity Developed, Low Intensity Developed, Cultivated, Grassland, Shrub Land, Shrub Wetland, Emergent Wetland, Shore Land, Bare Land, Water, Tundra, and Palustrine Aquatic Bed, as well as snow, clouds, and image background.

The image processing technique employed by Earthsat starts by classifying the two scenes using a combination of unsupervised and supervised methods. Unsupervised classification was used to create a signature file for 233 classes. The signature file was then run through a maximum likelihood supervised classification process. For a general description of these classification methods, see Jensen (1996). The resulting clusters were labeled using the Earthsat-developed addition to ERDAS Imagine 8.5 image processing software, called Geotools, which uses field and aerial photo-derived estimates. A change map was then constructed by using an Earthsat change detection technique called Cross Correlation Analysis (CCA). This analysis technique uses the labeled cluster file of the early-date image combined with the late-date multispectral image and statistically analyzes it against the labeled cluster file of the late-date image with the early-date multispectral image. Each pixel is ultimately placed into a change category (including "no change") based on the CCA process. More details on the method are located in the map metadata found in the C-CAP citation. The advantage is that it performs well regardless of seasonal differences because it uses former class boundaries summarized with new class signatures to determine the relationship between pixel brightness values and a feature class. In fact, direct pixel value comparison between the different scenes is not required (NOAA 2002).

The percentage area of forest land lost to urbanization and other land cover based on the NOAA change map was compared to the FIA estimates of percentage forest land change from 1980 to 1993 in the three New York FIA survey units that are wholly contained within the NOAA mapped area.

A change detection map based on the difference between pixel brightness values between two dates of Landsat imagery was also evaluated. For two counties in New Jersey, Monmouth and Ocean, change maps were produced using 1991 and 1997 Landsat images. Three different change layers were used to construct the map: a red band subtraction layer; a Normalized Difference Vegetation Index (NDVI) subtraction layer; and a layer consisting of the fourth principal component of a principal component analysis of the "brightness" and "greenness" layers of the *Tasseled Cap* transformation of both images. Again, for a detailed discussion of these bands, the NDVI, and band transformations, see Jensen (1996). In each of these layers, the pixel brightness values are highly correlated with a gain or loss of green biomass.

When the red spectral layer of the 1991 image is subtracted from the 1997 image, a high (bright) pixel value indicates a loss of forest canopy if the pixel area was forested in 1991. Green biomass is dark in the red band due to photosynthetic

Figure 1.—*Landsat images used for NOAA's Great Lakes Coastal Change Map (outlined square image footprints) and the area extent of the Change Map (dark gray). Note that two-thirds of New York was mapped. (Reproduced with permission from NOAA and Earthsat, Corp.)*

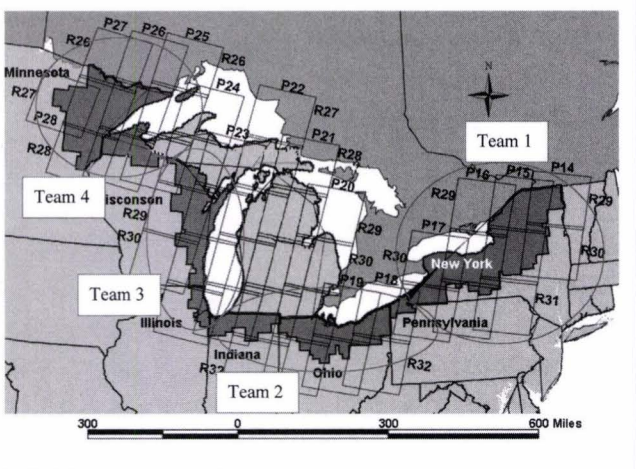
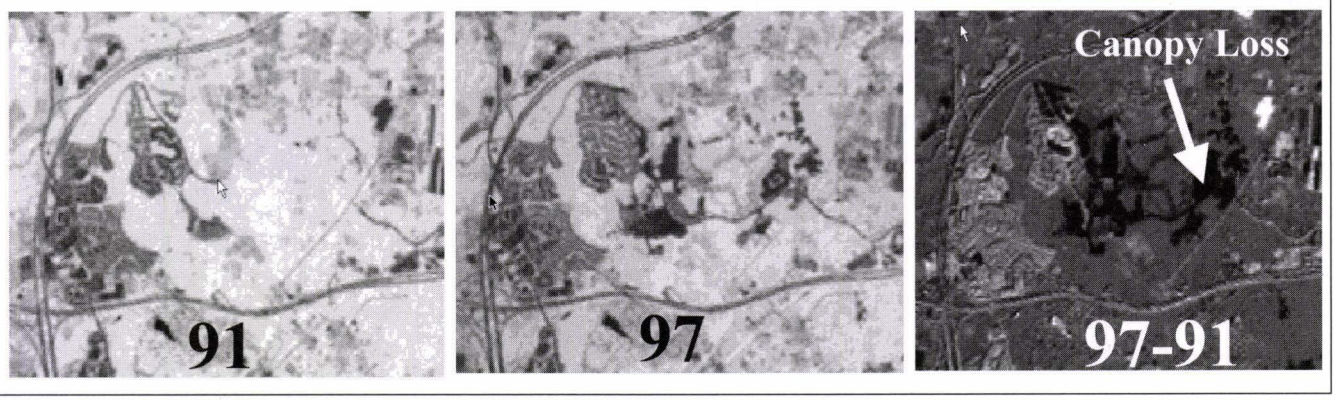


Figure 2.—*Urban development has removed forest cover between 1991 and 1997, as shown in these portions of NDVI scenes from the two dates. When the brightness values of the 1991 scene are subtracted from those of the 1997 scene, a forest canopy loss map is produced.*



light absorption. Similarly, when the 1991 NDVI is subtracted from the 1997 NDVI, a low (dark) pixel value indicates a loss of forest canopy. Green biomass has a high NDVI pixel value. A *Tasseled Cap* transformation produced a "brightness" band that is negatively correlated with forest canopies and a "greenness" band that is positively correlated with forest canopies for each of the images. Principal component analysis was applied to the four input layers to condense them into one biomass loss layer (Gessler *et al.* 1998). For each of the change images, a change map was constructed by selecting a pixel brightness threshold where forest canopy loss was on one side or the other. Threshold selection was based on aerial photo interpretation of areas where the forest canopy was completely removed (fig. 2).

Each of the change maps was very sensitive to even slight changes in the brightness value selected for the threshold. Furthermore, even though the maps were similar, the total number of pixels classified as "forest canopy loss" varied as much as 20 percent. So, a final forest canopy loss map was constructed by combining the three input maps. A pixel classified as "forested" in the USGS-MRLC map in 1991 and classified as "forest canopy loss" on all three change maps for New Jersey was required for a pixel to be labeled "forest canopy loss" in the final map.

Results

Binomial Probabilities

In our mathematical evaluation of how well FIA plots can estimate small amounts of forest loss equal to 5 percent of the total area, we found that the probability of a plot hitting an area of change exactly 5 times is 18 percent. In other words, 100 plots, distributed randomly at one 1-acre plot per 6,000 acres, will correctly estimate 5 percent loss of forest cover about one-fifth of the time. Only 1 percent of the time will no change of forest cover be detected at all; however, 50 percent of the time the estimate of forest loss will differ from that estimated by more that 1 percentage point, or by 9 square miles or more (fig. 3).

Figure 3.—*The probability that exactly a given number of plots will fall on change, if 5 percent of the area has changed and there are 100 plots. Note there is a relatively high probability (18 percent) the plots will estimate the exact area of change.*

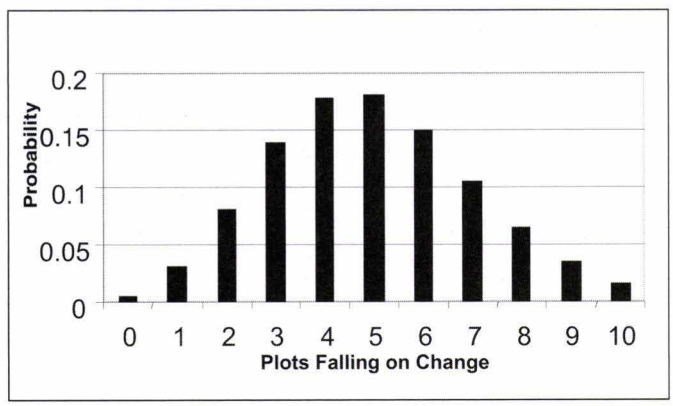
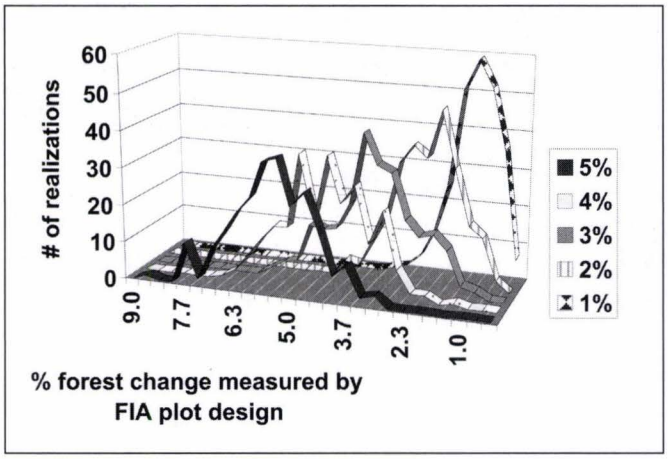


Figure 4.—*Histograms of 250 realizations each for a 1% through 5% GIS removal of the mapped forest cover of Connecticut (removed, at random, in unit areas of 7.5 acres). For each actual amount of forest removed, the frequency of realizations for each forest change percentage estimated by 300 forested plots is shown.*



GIS Experiment Using FIA Plot Locations in Connecticut and a Forest Cover Map of the State

In our GIS experiment, we found that the expected values of estimates of forest change are equal to the modes of the frequency distribution of 250 realizations for each actual percentage forest cover removed. Remember that a "realization" is the estimate of forest cover loss provided by 300 forested plots in Connecticut after a selected per-centage forest area is removed in multiples of 7.5-acre units. The range of estimates for each of the actual percentage removals was nearly identical to what the 99 percent confidence interval would be for a sample size of 300 and a fraction observed of 1 percent, 2 percent,...,5 percent forest cover loss (based on table 1.4.1 "Confidence intervals for binomial distribution" of Snedecor and Cochran (1967)) (fig. 4). Keep in mind that a 1 percent error in estimating change is equal to 29 square miles of forest land in the state of Connecticut. Before we leave the results of plot-based estimates of forest loss, consider that as the number of plots decrease, the ability to estimate small change is diminished. In a county like New Haven, CT, with 180,000 acres of forest land, a 1-percent loss would not be detected 75 percent of the time due to the low number (30) of forested plots.

Comparison of the NOAA/Earthsat Change Map with FIA Estimates

A summary of the forested Landsat pixels that changed in the 5-year period from 1995 to 2000 according to the NOAA/Earthsat change map of New York indicates that out of 15,817 square miles of forest, 251 square miles (1.5 percent) were lost to other land cover types. The average size of change was 1 acre with a range of 1/4 acre to 8,000 acres. Most of the change ranged from 1 to 38 acres. It is difficult to compare these results with forest change based on FIA plot data because the last periodic FIA survey of New York by the Forest Service was in 1993. The three northern FIA survey units in New York are contained within the NOAA mapped area and cover 67 percent of the acreage. Between 1980 and 1993 this area had a net increase of 143,000 acres of forest land, for a gain of 1.8 percent or 210 square miles (Alerich and Drake 1995).

In both the satellite-based estimate and FIA plot-based estimate, the map error and sampling error, respectively, are about the same magnitude as the change. The error matrix provided with the NOAA/Earthsat map estimates that the classification accuracy of forest land on the year 2000 map is 95 percent (96 percent correct omission error and 95 percent correct commission error) based on 750 ground truth plots, of which 200 plots were on forest land. This is a high level of accuracy not often exceeded in Landsat image based maps. Even though the forest loss estimate is smaller than the map error rate for forest land, experience shows that certain types of change should be accurately depicted, such as forest land to residential housing, commercial, and industrial areas. The NOAA map indicates that 16 square miles (0.01 percent) were lost to urbanization. Furthermore, the map shows the location and size of likely forest-loss patches.

Comparison of Forest Change Estimation: FIA Plots vs. Landsat Pixel Algebra Maps

Our forest loss map, based on an algebraic comparison of pixel brightness values of two Landsat scenes, shows a forest canopy loss of 3,547 acres in Monmouth County, New Jersey, between 1991 and 1997. The corresponding FIA estimate shows a net loss of 7,000 acres out of 90,300 acres of forest land from 1987 to 1999 (Griffith and Widmann 2001). If the rate of forest loss was evenly distributed over each of the 12 years, these two

estimates are in close agreement. Based on aerial photo comparison and the complete classification redundancy of the combined three image inputs to the satellite map, it is highly likely that the estimates of "loss of forest canopy" are accurate even though accuracy was not formally assessed. This method of classification does not necessarily allow a final estimate of how much forest land is converted to other land cover types because the area of lost forest canopy may regenerate back into a closed canopy forest, as it might after fire or harvesting.

In adjacent Ocean County, New Jersey, the Landsat map estimated a forest canopy loss of 2,047 acres (1991–1997). FIA estimates a 42,000-acre gain (1987–1999), up from a forested area of 204,000 acres in 1987. Since FIA reports only net forest change, the 2,047 acres of possible forest loss would be difficult to detect. Furthermore, the FIA estimate has a sampling error of about 8 percent, so sampling error alone would not allow for accurate detection of a 2,000-acre loss of forest land.

Conclusions

The authors reject the hypothesis that the density of FIA plots, 1 plot per 6,000 acres, does not permit accurate estimates of forest land loss if the area lost is low—1 to 5 percent of the total forest area. This rejection, however, requires that the number of plots is reasonably large, which in the FIA survey program also means a reasonably large survey area. It does seem paradoxical that a sample that covers only 0.017 percent of the area can detect and accurately estimate a small amount of change, but one of the properties of statistics is that it only requires there be a large absolute number of plots, not a large sample relative to the population. However, in small counties that have few plots, accurate estimates of forest loss are not possible. This also limits the ability of FIA plots to provide useful information on the spatial distribution of forest loss.

So, accurate FIA plot-based state estimates of forest loss are possible, even if the percentage loss is low. Of course, if the estimate is only of net forest change, the important measurement of forest loss is not likely to be obtained. In sufficiently large areas, such as at the State level, the FIA program could add forest area loss to its list of reported estimates. Landsat-derived forest change maps can provide both quantitative and spatial information on how, where, and why forest land is lost. Map error and the inability of moderate resolution imagery to classify seedling/sapling covered forest areas will continue to degrade the accurate interpretation of these forest cover maps. However, satellite imagery can be an efficient tool for small and large area forest assessments, especially in those frequently found cases in the East where forest land is lost to urbanization, roads and freeways, and reservoirs.

The NOAA/Earthsat change map seems to be a high-quality map asset for analyzing landscape changes. State and Federal land management agencies may wish to consider using it to provide large area land cover change information. The authors find, as many have before, that Landsat can be used quickly, simply, and robustly to detect forest area loss.

The USDA Forest Service Forest Inventory and Analysis program will continue to explore the use of remote sensing to augment the ability to determine the extent, condition, and trends of forest land.

Acknowledgments

The authors wish to thank Greg Koeln, Eleanore Meredith, Dave Cunningham, and Francois Smith of Earthsat Corp. and Mark Finkbeiner of NOAA for providing the NOAA/Earthsat change maps of New York and helping us evaluate their contents.

Literature Cited

Alerich, C.; Drake, D. 1995. Forest statistics for New York: 1980 and 1993. Resour. Bull. NE-132. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 249 p.

Brooks, R.; Kitteredge, D.; Alerich, C. 1993. Forest resources of southern New England. Resour. Bull. NE-127. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 71 p.

Gessler, P.; Grey, D.; Hoppus, M. 1998. Change detection using remotely-sensed data: a USDA Forest Service training course manual. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center. 50 p.

Griffith, D.; Widmann, R. 2001. Forest statistics for New Jersey: 1987 and 1999. Resour. Bull. NE-152. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 70 p.

Hansen, M.; Frieswyk, T.; Glover, J.; Kelly, J. 1992. The east-wide forest inventory data base: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 48 p.

Huntsburger, David; Billingsley, Patrick. 1977. Elements of statistical inference. Boston, MA: Allyn and Bacon, Inc. 385 p.

Jensen, J.R. 1996. Introductory digital image processing. Upper Saddle River, NJ: Prentice Hall. 318 p.

NOAA Coastal Services Center/Coastal Change Analysis Program (C-CAP). 2002. C-CAP Great Lakes 1995-2000 land cover metadata. Charleston, SC: NOAA Coastal Change Analysis Program (C-CAP). Available at: http://www.csc.noaa.gov.

Snedecor, George W.; Cochran, William G. 1967. Statistical methods. 6th ed. Ames, IA: Iowa State University Press.

Spatially Locating FIA Plots from Pixel Values

Greg C. Liknes[1], Geoffrey R. Holden[1], Mark D. Nelson[1], and Ronald E. McRoberts[2]

Abstract.—The USDA Forest Service Forest Inventory and Analysis (FIA) program is required to ensure the confidentiality of the geographic locations of plots. To accommodate user requests for data without releasing actual plot coordinates, FIA creates overlays of plot locations on various geospatial data, including satellite imagery. Methods for reporting pixel values associated with FIA plots that reduce the likelihood of inadvertent release of plot locations were tested. The number of plots that can be correctly located using only pixel values was reduced by perturbing image band values, averaging band values of neighboring pixels, and reducing the number of image bands.

In the FY2000 Consolidated Appropriations Bill (PL 106-113), Congress included language that modified the Food Security Act of 1985 (7 U.S.C. 2276(d)) to add FIA data collection to a list of items requiring confidential treatment. As a result, the FIA program must ensure the confidentiality of coordinates of field plots, maintaining sample integrity and protecting the privacy of landowners granting FIA access to their lands.

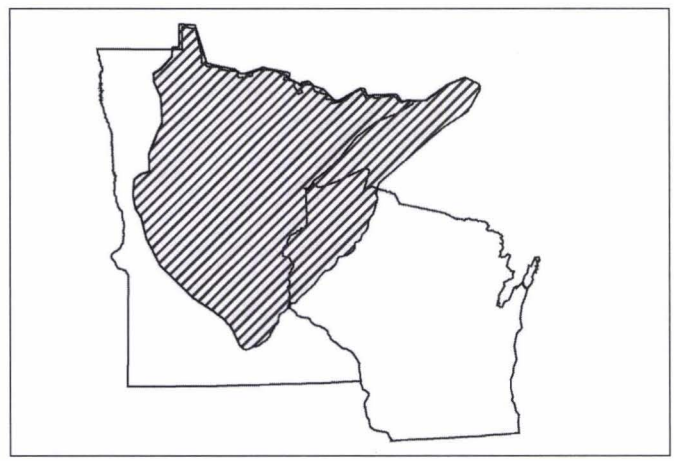
Researchers interested in using FIA plot data as a source of training data for remote sensing classification do not necessarily need plot location information. Rather, FIA staff could extract spectral values from satellite imagery at plot locations and provide the spectral information along with requested FIA plot data (e.g., percent forest cover, basal area), but without plot coordinates. This method of providing FIA data, however, may inadvertently reveal plot location information when a pixel associated with a plot has a combination of spectral values unique to that pixel. In this study, we examined the probability of determining a plot's location from spectral information.

Data

We used data from the Multi-Resolution Land Characterization (MRLC) Consortium National Land Cover Dataset 2000 (NLCD 2000) mapping zone 41 (Homer and Gallant 2001). Zone 41 encompasses part of Minnesota, a portion of western Wisconsin, and Isle Royale in Michigan (fig. 1). The data set is comprised of multiple Landsat Thematic Mapper (TM) and Landsat Enhanced Thematic Mapper+ (ETM+) scenes at 30-m pixel resolution.

Thermal bands and Landsat-derived data, including tasseled cap transformations (greenness, soil brightness, and wetness), and textural information, were available for the area of interest. Additionally, digital elevation data and State Soil Geographic (STATSGO) (http://www.ftw.nrcs.usda.gov/stat_data.html) soil data (available water content, carbon content, quality) were available, as well as a data band containing the date of the input image used for each pixel. In total, 24 bands of data (19 continuous and 5 categorical) were used in this study (table 1). Of the 5,939 measured FIA plots falling within zone 41 cloud-free areas, 200 were randomly selected for testing.

Figure 1.—*MRLC NLCD 2000 mapping zone 41 (hatched area) covering a large portion of Minnesota, some of Wisconsin, and Isle Royale in Michigan.*

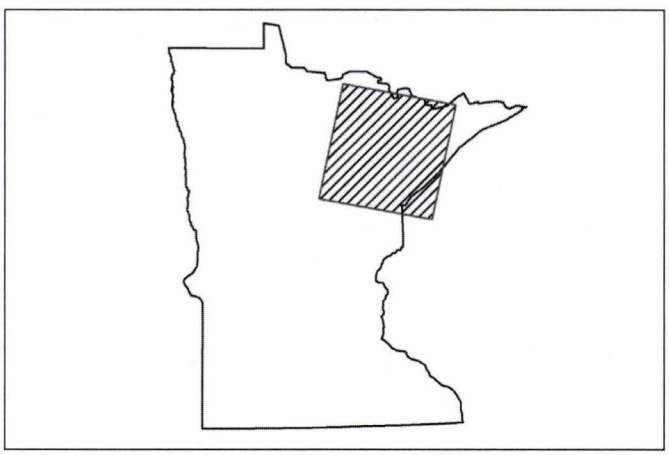


[1] GIS/RS Specialist, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108.
[2] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108.

Table 1.—*Available data bands and their data types for MRLC NLCD 2000 mapping zone 41*

Data band	Band name	Data type
1	Texture (band 1—leaf-on)	continuous
2	Texture (band 4—leaf-on	continuous
3	Texture (band 7—leaf-on)	continuous
4	Greenness (spring)	continuous
5	Brightness (spring)	continuous
6	Wetness (spring)	continuous
7	Greenness (leaf-on)	continuous
8	Brightness (leaf-on)	continuous
9	Wetness (leaf-on)	continuous
10	Greenness (leaf-off)	continuous
11	Brightness (leaf-off)	continuous
12	Wetness (leaf-off)	continuous
13	Thermal (spring)	continuous
14	Thermal (leaf-on)	continuous
15	Thermal (leaf-off)	continuous
16	Elevation	continuous
17	Slope	continuous
18	Aspect	categorical
19	Soil quality	categorical
20	Soil carbon	continuous
21	Soil available water content	continuous
22	Spring date	categorical
23	Leaf-on date	categorical
24	Leaf-off date	categorical

Figure 2.—*Location of Landsat path 27 row 27 (P27R27), Minnesota.*

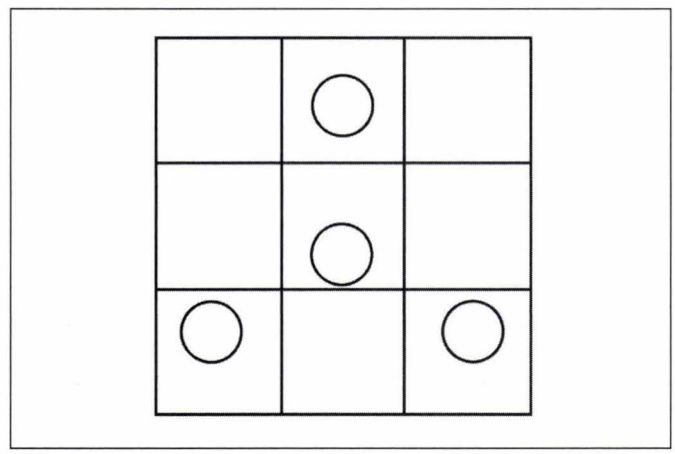


A second study was performed using the data from a single Landsat ETM+ scene, path 27 row 27 (P27R27) (fig. 2), which is one of the scenes contained in zone 41. For this single scene, imagery from three different dates (31 May 2000, 5 July 2001, and 5 November 1999) was available. The six reflective bands (1-5, and 7) for each date were used, totaling 18 bands for this scene. Of the 1,277 measured FIA plots falling within the scene, 200 were randomly selected for testing.

Methods

To determine if providing spectral information reveals plot location information, pixel values at each of the 200 plot locations were extracted for all the data bands in each study. Each field plot consists of four 7.31-m (24-foot) radius circular subplots. The subplots are configured as a central subplot and three peripheral subplots with centers located 36.58 m (120 ft) and azimuths of 0°, 120°, 240° from the center of the central subplot. The plot/pixel arrangement shown in figure 3 is the most probable arrangement and therefore was used in this study. The arrangement of pixels relative to subplots can change slightly depending on where the center subplot falls within a pixel. Three methods for extracting pixel values were considered: (1) values associated with the pixel containing the central subplot, (2) the average of the pixel values associated with the four subplots (fig. 3), and (3) the average of the pixel values in a 3x3 window centered on the pixel associated with the center subplot. For methods 2 and 3, new images were generated using

Figure 3.—*Location of FIA subplots relative to Landsat TM 30-m pixels.*

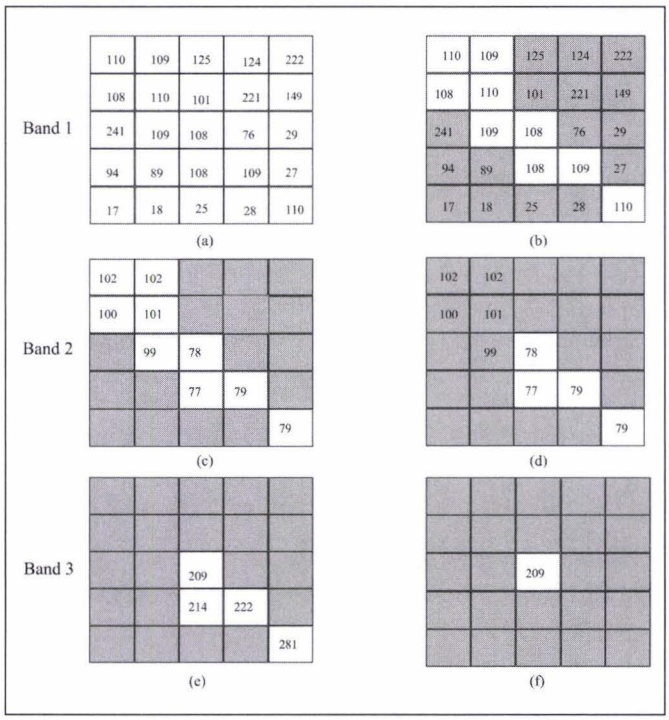


focal analysis techniques, with the averaged values applied at every pixel.

To test the effect of alterations of pixel values on a user's ability to determine correct and unique plot locations, extracted values were perturbed for each of the three methods by an integer randomly selected from the following intervals: [0], [-1, 1], [-2, 2], and [-3, 3]. The combination of the four perturbation intervals with the three methods resulted in 12 tests for both the zone 41 and the P27R27 studies. Additionally, the 12 tests were repeated for a single date (leaf-on) of P27R27.

For each test, the extracted values for each plot were compared to all image pixel values in respective bands. To illustrate, consider a 25-pixel, 3-band image (fig. 4), and suppose band values were extracted at the location of a plot falling within the image. Values for bands 1, 2, and 3 were 108, 79, and 209, respectively. Furthermore, a perturbation of [-1, 1] was then applied to the extracted values, resulting in a set of band values of 109, 78, and 209. The data from band 1 are shown in figure 4(a). All pixels with values not within ±1 of the perturbed value (109) are grayed out in figure 4(b). The data from band 2 are displayed in figure 4(c). Note that all gray pixels no longer need to be checked against the perturbed values because they failed to match the band 1 perturbed value. The remaining data that fell within ±1 of the band 2 perturbed value (78) are shown in white in figure 4(d), and the non-matching pixels are now shown in gray. The same process is applied for band 3 data (perturbed value 209) in figures 4(e) and 4(f). In the end, a single pixel remains that matches all three perturbed values. If Landsat TM pixels are represented in figure 4, this process reveals the location of an FIA plot to within 30 m (the resolution of a Landsat pixel), assuming the true ground position of each pixel is known.

Figure 4a–f.—*The sequence of steps involved in determining from which pixel an extracted set of values came.* Extracted values with [-1, 1] perturbation applied: band 1 = 109, band 2 = 78, band 3 = 209. *In this example, pixels are grayed out if they don't fall within ±1 of the perturbed values for the corresponding band. Grayed pixels are then ignored for each subsequent band.*



For each test, if only a single pixel matched the extracted values for all bands, its location was compared with the FIA plot locations to determine if the plot had been correctly "located" using only spectral information. All processing was performed using ESRI Grids in ArcInfo and automated using an Arc Macro Language (AML) script.

Results and Discussion

For zone 41, all of the 200 FIA plots were correctly located using only spectral (pixel) information for the single pixel and 4-subplot average methods with no perturbation, while 98 percent were located for the 3x3 average method (table 2). As the amount of perturbation increased, the percentage of plots correctly located decreased. This is because the number of potential matches for each pixel increased as the interval around each extracted value increased, thereby reducing the likelihood of uniquely identifying a pixel. Although the percentage of plots correctly located decreased with increased perturbation, more than half of the plots were located for each of the three methods at all perturbation levels.

For the P27R27/multiple date study (18 bands), all of the 200 FIA plots were correctly located using only spectral (pixel) information for each of the three methods with no perturbation (table 3). As was the case for zone 41, the percentage of plots located in P27R27 decreased as the amount of perturbation increased. For P27R27, the percentage of plots located is not

Table 2.—*Percentage of FIA plots located from MRLC NLCD 2000 mapping zone 41 pixel information for various methods and perturbation levels*

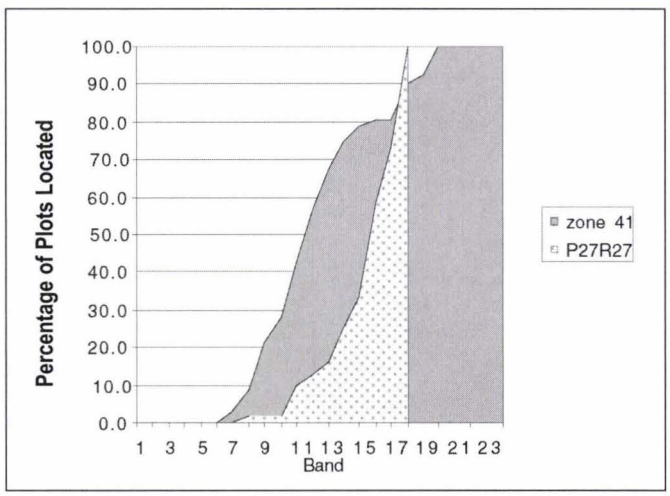
Method	Perturbation	[0]	[-1,1]	[-2,2]	[-3,3]
Single pixel / center subplot		100	94	89	72
3x3 average		98	88	72	53
Four subplot average		100	91	76	58

Table 3.—*Percentage of FIA plots located from Landsat TM path 27 row 27 pixel information for various methods and perturbation levels*

Method	Perturbation	[0]	[-1,1]	[-2,2]	[-3,3]
Single pixel / center subplot (three dates)		100	97	63	32
3x3 average (three dates)		100	71	31	14
Four subplot average (three dates)		100	84	41	19
Single pixel / center subplot (summer only)		17	0	0	0
3x3 average (summer only)		6	1	0	0
Four subplot average (summer only)		7	0	0	0

nearly as high as in zone 41 for the [-3,3] perturbation. In figure 5, the cumulative percentage of extracted values that identify single pixels (single pixel method, [-3,3] perturbation) is graphed against the band that has been processed. In both the zone 41 and P27R27 tests, a high percentage of the plots were located before all bands were processed. Thus, the difference is not simply a matter of the number of bands (18 vs. 24). All zone 41 and P27R27 bands were 8-bit data (256 possible values), with the exception of elevation (zone 41 only), which had values ranging from 177 to 701. The greater variability in this band may have contributed to the higher percentage of plots located in zone 41 in the single pixel method, [-3,3] perturbation test. Also, the elevation band was processed first for all zone 41 tests, which may have significantly impacted at which step unique pixels were identified in the process (fig. 5).

Figure 5.—*Cumulative percentage of FIA subplots correctly located as a function of sequentially adding bands, MRLC NLCD 2000 mapping zone 41 (24 bands) and Landsat TM path 27 row 27 (18 bands).*



For the P27R27/single date study (6 bands), few plots were correctly located. For example, 17 percent of plots were correctly located in the single pixel/no perturbation test. In this case, the noticeable decrease in correctly located plots is most likely due to the small number of bands considered. In both the zone 41 study (24 bands) and the P27R27/multiple date study (18 bands), most plots could not be found until after the sixth band (fig. 5).

Conclusions

If the method for extracting pixel values at plot locations is known and no perturbation is applied to the pixel values, it is possible to locate correctly all or nearly all of the plots with 18 or 24 bands of data. For a single-date, 6-band image, few plots can be correctly located if no perturbation is applied, and

almost no plots if a perturbation is applied. Increasing the perturbation reduces the number of plots that can be correctly located but also reduces the usefulness of the pixel information for some users' applications.

In the future, if spectral data are requested for FIA plot locations, the method described in this paper could be used to screen plots for possible inadvertent disclosure of plot information. Currently, the greatest drawback to the procedure is the time involved in processing an array of extracted values. Future work may include optimizing this processing or develop alternative processing procedures.

Literature Cited

Homer, C.G.; Gallant, A. 2001. Partitioning the conterminous United States in mapping zones for Landsat TM land cover mapping. USGS Draft White Paper.

Variable Selection Strategies for Small-area Estimation Using FIA Plots and Remotely Sensed Data

Andrew Lister, Rachel Riemann, Jim Westfall, and Mike Hoppus[1]

Abstract.—The USDA Forest Service's Forest Inventory and Analysis (FIA) unit maintains a network of tens of thousands of georeferenced forest inventory plots distributed across the United States. Data collected on these plots include direct measurements of tree diameter and height and other variables. We present a technique by which FIA plot data and coregistered remotely sensed raster data were used to predict the basal area of deciduous trees at a spatial resolution of 30 m. Results varied, generally indicating that culling putatively unrelated variables did not improve estimates over those obtained using all the potential variables in the model.

The USDA Forest Service's Northeastern Forest Inventory and Analysis unit (NE-FIA) is charged with conducting a portion of a national forest inventory. NE-FIA uses data collected on a network of ground plots to produce reports on the status of the region's forests.

In addition to tabular reports, analysts and data consumers frequently request spatially explicit, highly resolute maps of forest variables. To produce these maps, data from geographic information systems (GIS) and satellites are often used to build models that predict attributes such as volume, biomass, and basal area.

There are many choices of GIS data and satellite layers for a given region. The National Land Cover Dataset (NLCD) project, a USGS-led, collaborative effort among several governmental and nongovernmental groups, is producing national land cover maps using GIS and satellite data. The USGS Eros Data Center compiled 18 GIS and satellite imagery layers for a mapping area covering several Mid-Atlantic States (NLCD mapping zone 60).[2] These data layers are coregistered, so they can be easily combined with NE-FIA plot data to produce a data set that can be used for predictive modeling. The goal of the current study was to assess the effects of subsetting these 18 layers to arrive at a model training set that would lead to more accurate predictions of the basal area of deciduous trees.

Methods

The study area included that portion of New Jersey covered by the NLCD imagery data (fig. 1). Data were collected on NE-FIA plots in New Jersey between 1998 and 1999.[3] The total amount of deciduous tree basal area measured on each plot was used as the dependent variable in the predictive modeling. Only

Figure 1.—*The study area in central and southern New Jersey; 141 homogeneous, forested plots were used for the analysis.*

[1] Forester, Research Forester, Research Forester, and Supervisory Forester, respectively. U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA 19073. Phone: 610–557–4038; e-mail: alister@fs.fed.us.

[2] Homer, C.; Gallant, A. 2001. Partitioning the conterminous United States in mapping zones for Landsat TM land cover mapping. USGS Draft White Paper, on file at USGS Eros Data Center, 47914 252nd Street, Sioux Falls, SD 57198.

[3] USDA Forest Service. 2000. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.4. USDA Forest Service, internal report. On file at USDA Forest Service, Washington Office, Forest Inventory and Analysis, Washington, DC.

completely forested plots were used in the analysis.

The portion of the NLCD data in New Jersey was used as a source of potential predictor variables. The NLCD data set was assembled by mosaicking, georeferencing, and radiometrically correcting three-season satellite imagery collected by the Landsat 7 satellite between 1999 and 2001 (USGS Eros Data Center 2002). These assembled raw images were transformed using the Tassled Cap (TC) transformation, a procedure that produces new images consisting of three layers per original seasonal six-band image (USGS Eros Data Center 2002). The TC transformation typically is used because the composite layers have a higher correlation with some features of vegetation than do the constituent layers. In addition to these nine TC layers, elevation, slope percentage, aspect, and slope position index were derived from digital elevation models (DEMS), which are raster GIS layers with a value for elevation at each pixel location. Slope, aspect, and slope position (ranging from 0 in the valley to 100 on the ridgetop) also were calculated for each pixel using a GIS. Soil quality, available water content (awc), and soil carbon percentage (variables that often are considered when measuring site quality) were derived from the STATSGO soils data set produced by NRCS (USDA Soil Conservation Service 1993). Layers consisting of geographic Easting and Northing also were created. All NLCD data layers were coregistered, standardized to be within the range of 0-255, and resampled to a 30-m pixel size.

Values of predictor variables at plot locations were obtained with Erdas Imagine software. Scatterplots of deciduous basal area vs. each of the predictor variables were generated and correlation matrices were created with SAS software. To create a subset of predictors for modeling, variables that were not significantly correlated with basal area were excluded from modeling, as were plots that were significantly correlated but subjectively considered weakly related after assessing the scatterplots.

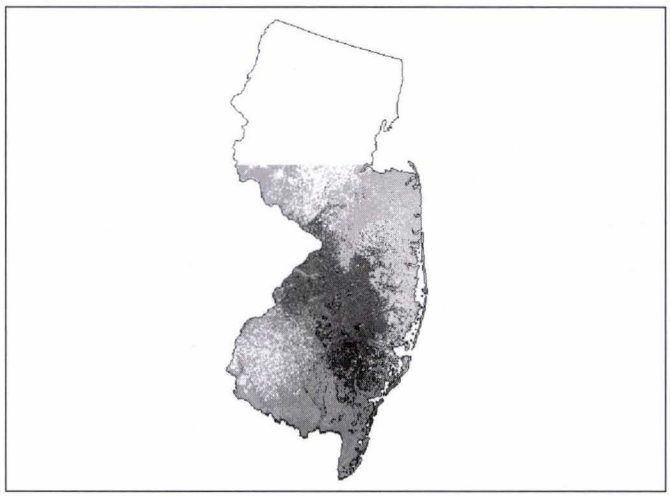
The two modeling data sets (the full set and the subset) were used to produce maps of deciduous basal area for each 30-m pixel defined by the predictor layers. The technique used was a minimum-distance supervised classification, which is in effect a *k*-nearest neighbor imputation with a *k* of 1 (McRoberts *et al.* 2002, Franco-Lopez *et al.* 2001). This procedure is based on the multidimensional Euclidean distance between pixels where basal area is unknown but the predictor variables have known values, and a pixel with known values for both basal area and predictor variables. The basal area of the plot whose associated pixel has the smallest multidimensional Euclidean distance from the unknown pixel is assigned to the pixel being evaluated. Each pixel is treated in this way until a continuous map of basal area is produced.

The modeling procedure is such that a given plot's value never influences the prediction at its own location; it always is a different plot whose value is assigned to the pixel on which a plot sits, making it possible to use the modeling data for validation. To assess the accuracy of the resulting maps, scatterplots of observed vs. predicted basal area were generated from the original data, and simple linear regression models describing the relationship between observed and predicted values were created. Histograms of absolute error were generated for both maps.

Table 1.—*Correlation coefficients (r values) and p values from correlation analyses of the relationship between deciduous basal area and several predictor variables (N=141)*

Predictor variable	r value	p value
Position index	– 0.09	0.31
Slope	0.29	<0.001
Aspect	0.26	<0.01
Elevation	0.25	<0.01
Easting	– 0.54	<0.0001
Northing	– 0.24	<0.01
Soil water content	0.44	<0.0001
Soil carbon	0.08	0.33
Soil quality	0.33	<0.0001
Summer brightness	0.60	<0.0001
Summer greenness	0.65	<0.0001
Summer wetness	– 0.10	0.24
Fall brightness	0.58	<0.0001
Fall greenness	0.59	<0.0001
Fall wetness	– 0.06	0.45
Spring brightness	0.30	<0.01
Spring greenness	– 0.54	<0.0001
Spring wetness	– 0.41	<0.0001

Figure 2.—*Final map of predictions of deciduous basal area for central and southern New Jersey. The map was produced using all 18 GIS and imagery layers. Lighter values indicate higher levels.*

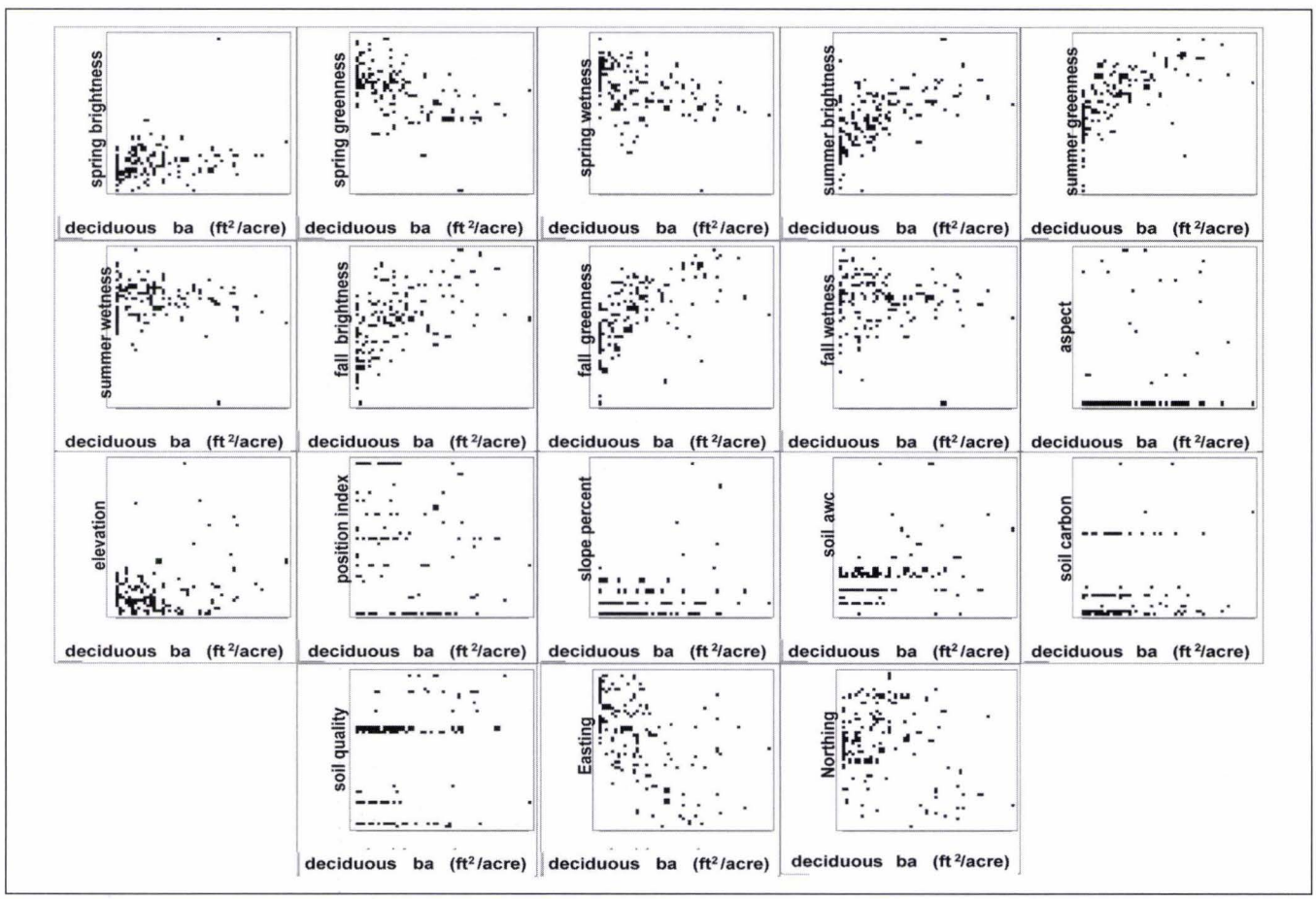


Results and Discussion

The final map is shown in figure 2. The correlation statistics and resulting p values are shown in table 1. Fourteen of the 18 original predictor variables had significant correlation coefficients. Position index, soil carbon, and summer and fall wetness (two of the TC layers) were not significant ($p>0.05$). After subjectively assessing the scatterplot matrix (fig. 3), we decided to eliminate aspect, elevation, slope percentage, and Northing.

We had hypothesized that the DEM-based layers would be related to site quality and probability of development; low slope sites closer to a valley floor should have deeper, moister soil and be more prone to human development than sites on steep hillsides or ridgetops. There were no or only weak relationships between topographic site factors and basal area of

Figure 3.—*Scatterplot matrix showing relationship between the basal area of deciduous trees on FIA plots and several GIS and imagery-based values (see Methods for information on predictor variables). Y axis values ranged from 0 to 255; x axis values ranged from 0 to 192 ft²/acre (N=141).*



deciduous trees measured on NE-FIA plots in our study area, possibly because there are higher order interactions among topographic variables or between them and other unmeasured variables. Similarly, Xu and Prisley (2000) hypothesized that the local variation in carbon distribution could be due to factors such as previous land use, forest growth phase, forest type differences, geomorphology, natural disasters, or other unmeasured factors.

The lack of relationship between basal area and soil carbon might be caused by the same phenomenon or by the nature of the soil carbon data. Soil carbon may not be measurably biologically related to basal area production. The relationship between Northing and basal area did not appear to be linear. Perhaps the relationships could have been improved via transformation or by creating composite variables to test the effects of interactions, but we chose not to transform the data to preserve biological interpretability of our model outputs.

The TC wetness layer historically has been used to represent different levels of soil moisture. Perhaps during summer and fall, little bare soil was exposed on the FIA plots, making the wetness layer less useful during these seasons. However, before the deciduous trees produce leaves in spring, the satellite acquires reflected light from bare soil beneath the trees and thus might be measuring an ecological factor that affects deciduous basal area.

The scatterplots and diagnostic statistics of the regressions of observed vs. predicted for the full model and for the subset model are shown in figures 4a and 4b, respectively. Considering the shape of the scatterplot and the regression outputs, the full model performed better than the subset model. The R^2 value was higher, the slope was closer to 1, and the y intercept was closer to zero (figs. 4a - 4b). The histogram of absolute errors (fig. 5) indicates that the subset model performed slightly better in the second and third lowest basal area categories, but the subset model's errors generally had higher variances than those from the full model.

These results were unexpected. We had hypothesized that several of the potential predictor layers would be extraneous, that is, the effects of strong predictors on the estimate would be diluted. But we found that the full model performed better in a validation. The scatterplots indicate that the full model's accuracy was consistent throughout the distribution of observed val-

Figure 4.—*Observed vs. predicted scatterplots from validation of the full model (A) and the subset model (B). The full model used 18 GIS and imagery layers; the subset model used only 10 (N=141).*

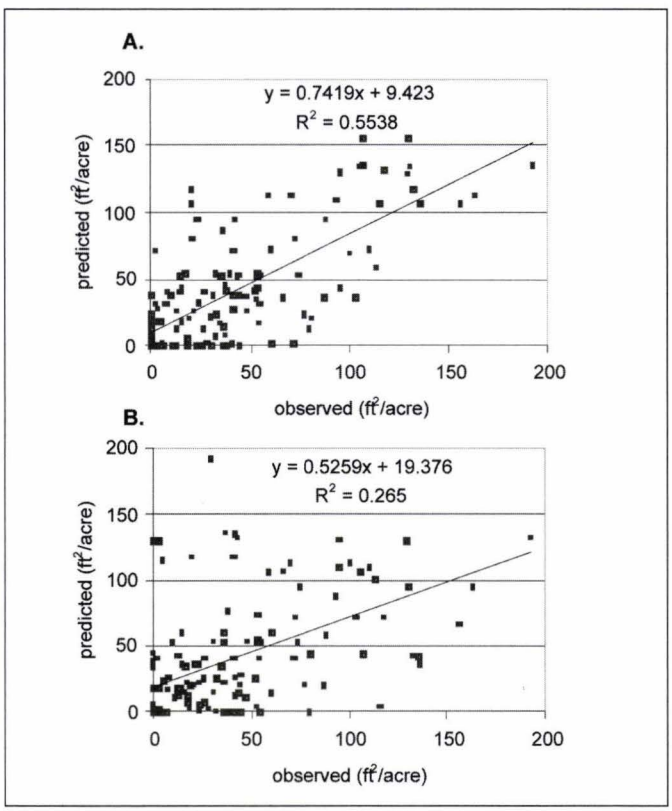
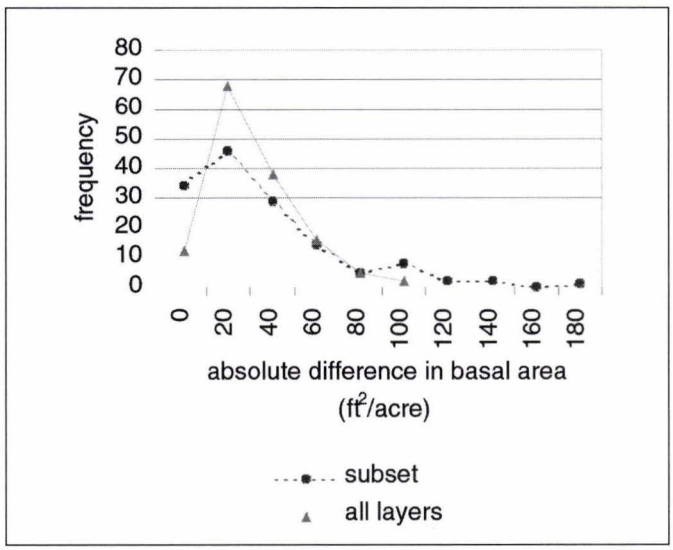


Figure 5.—*Absolute error histograms of the full model and the subset model. Absolute errors were calculated for each model by calculating the absolute difference between observed and expected (N=141).*



ues, whereas for the subset model, the accuracy was much worse in the lower tail of the observed data's distribution. A possible explanation is that, when extraneous data layers are used as predictors, the potential negative impact of anomalous model training data is mitigated. An unknown pixel that is incorrectly classified by a "bad" training site using the subset model might not be classified incorrectly if the distances are perturbed slightly by the addition of extraneous data layers. That extra distance raises the probability that a "better" training site might be assigned to that unknown location.

If this is the case, there must be a tradeoff between diluting the strength of mechanistic relationships between predictors and dependent data and susceptibility to poor training data. Our results suggest that the extraneous data layers served as a safety net, removing some of the effects of outlaying data points but not increasing the overall variance of our residual error to an unacceptable level.

In future studies, we plan to analyze the effects of individual training data sites on the accuracy of our modeling. Much of the variance in our absolute errors may be due to rogue training data. We also plan to test the effects of transforming the variables and creating composite layers consisting of interactions of the GIS and imagery layers. And we will investigate additional variable reduction methods, including multiple linear regression, principal components analysis, and other univariate and multivariate techniques. We also will assess the relationship between the amount of training data used and the number of spectral bands used. A phenomenon called the "Hughes Phenomenon" (Hughes 1968) occurs when accuracy is degraded when one increases the number of predictor variables but does not change the number of training sites.

Literature Cited

Franco-Lopez, H.; Ek, A.R.; Bauer, M.E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of the Environment. 77(3): 251–274.

Hughes, G.F. 1968. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory. 14: 55–63.

McRoberts, R.E.; Nelson, M.D.; Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors. Remote Sensing of the Environment. 82: 457–468.

U.S. Department of Agriculture, Soil Conservation Service. 1993. State soil geographic data base (Statsgo) data user's guide. Misc. Pub. 1492. Washington, DC: U.S. Department of Agriculture, Natural Resources Conservation Service.

U.S. Geological Survey, Eros Data Center. 2002. MRLC 2000 image preprocessing procedure. http://landcover.usgs.gov/pdf/image_preprocessing.pdf. and on file at USGS Eros Data Center, 47914 252nd Street, Sioux Falls, SD 57198.

Xu, Y.J.; Prisley, S.P. 2000. Linking STATSGO and FIA data for spatial analyses of land carbon densities. In: Hubbard, W.G; Jordin, J.B., eds. Proceedings of the 3rd USDA Forest Service Southern GIS Conference; 2000 October 10–12; Athens, GA.

Distributing FIA Information onto Segmented Landsat Thematic Mapper Images Stratified with Industrial Ground Data

Tripp Lowe[1], Chris J. Cieszewski[2], Michal Zasada[3,4], and Jarek Zawadzki[3,5]

Abstract.—The ability to evaluate the ecological and economical effects of proposed modifications to Georgia's best management practices is an important issue in the State. We have incorporated tabular FIA data with Landsat Thematic Mapper satellite images and other spatial data to model Georgia's forested land and assess the area, volume, age, and site quality of two stream and road buffer regimes. Each regime included different buffer widths for perennial and intermittent streams and slope classes. We discuss here the technical details of this work.

With ever-increasing public awareness of the welfare of natural resources, land managers are under great pressure to manage our forestlands with publicly acceptable stewardship. One of the prevalent issues is maintaining and improving the quality of our streams, public lakes, reservoirs, and wetlands. As a guideline for the forestry community, the Georgia Forestry Commission has compiled a set of "common sense, economical, and effective" methods, called "best management practices" (BMPs) (GFC 1999), designed to reduce nonpoint source pollution and protect the waters. While the BMPs are now voluntary in Georgia, many forest managers in the State believe that the regulations within the riparian zones should be further expanded.

For each change considered, we must be able to evaluate the possible ecological and economical impacts. As a basis for these evaluations, we need to know the amount and type of land that will be influenced by the considered change. Here we describe the use of currently available GIS and remote sensing technologies for evaluating such potential adjustments to the BMPs and their likely impacts on the resource availability. We performed several analyses of Georgia's timberlands in which tabular FIA data were incorporated into stands derived from various types of GIS data.

Objectives

The main objective of this study was to model and estimate the cover types in buffers around water resources and roads. This included spatially explicit analysis of various types of ground measurement, GIS, and other remote sensing data and modeling the distribution of forest inventory within the buffer cover type assignments. These analyses were necessary for estimating the effects of establishing different riparian zone buffers in terms of their total area, volume, age, and site quality at the State level as well as at a FIA-unit conglomerate.

Data

We analyzed various spatial data from several different sources. Vector hydrology and county boundary data sets, and raster National Elevation Datasets (NED) were downloaded from the Georgia GIS Data Clearinghouse located at http://www.gis.state.ga.us/Clearinghouse/clearinghouse.html. The hydrology data set includes rivers, streams, and artificial flow paths through water bodies captured from USGS 7.5-minute topographic maps. The NED data contain elevation information for the county at a resolution of 30 m.

The raster "1998 landcover map of Georgia" (GAP) data set, produced by the Natural Resource Spatial Analysis Laboratory (NARSAL, Institute of Ecology, University of Georgia), was used to quickly assess landcover types throughout the entire State. The GAP data set was generated by classifying Landsat Thematic Mapper (LTM5) images captured in

[1] GIS Analyst, [2] Assistant Professor, [3] Postdoctoral Fellow, respectively, Warnell School of Forest Resources, University of Georgia, Athens, GA.
[4] Assistant Professor at Faculty of Forestry, Warsaw Agricultural University, Poland.
[5] Assistant Professor, Environmental Engineering Department, Warsaw University of Technology, Poland.

the summers and winters of 1997 and 1998 and distinguishes among 18 different general landcover types. We evaluated areas classified as "clearcut/sparse," "deciduous forest," "evergreen forest," "mixed forest," and "forested wetland." All other areas classified as something else were ignored. Also obtained from NARSAL was the "Trout Streams of Georgia" vector data set including all streams in Georgia classified as habitable for trout.

Eight LTM 7 data sets were used as a second method to evaluate land cover throughout the State. The winter scenes were captured in November of 1999 and encompass the entire State except for the extreme northwestern and southern parts. LTM 7 satellite data capture information about the Earth's surface in three visible portions, two infrared, one thermal, and one panchromatic portion of the electromagnetic spectrum. Information from the visible and infrared bands was captured at a 30-m resolution, 60 m in the thermal band, and 15 m in the panchromatic band.

Methods

The analyses conducted in this project can be classified into five separate steps. The LTM images were classified by basal area using industrial ground data in the "image classification" phase. The images in the "image segmentation" phase were converted to homogenous polygons representing either coniferous, deciduous, or mixed timber stands. FIA information was distributed to those polygons using two different methods in phase 3, "FIA information distribution." Riparian and transportation buffers were generated in the "buffer creation" phase. Hydrology buffers were created using two buffer widths that incorporate three slope classes, and transportation buffers were generated using two buffer widths. In the final "data intersection" phase, the LTM-generated polygons and buffers were combined.

Image Classification

As a first-stage sample of landcover type, the GAP data set was used to locate coniferous and nonconiferous lands. Areas that were not classified by GAP as evergreen forests were masked out and ignored in subsequent analysis. The evergreen areas in the LTM images were classified by basal area using ground-measured industrial data recorded in 1998 (Ruefenacht *et al.* 2002) and the Euclidian spectral distance (Ruefenacht *et al.* 2002) from the "low basal area" signature (eq. 1):

$$\sqrt{\sum_{i=1}^{6}(S_i - A_i)^2}$$

where,

S = the seed pixel's LTM value

A = the adjacent pixel's LTM value

i = the LTM band (excluding the thermal band 6)

Due to the limited number of large, low basal area stands in the industry data set, we derived the "low basal area" signature by first generating a "high basal area" signature by sampling LTM pixels within those stand boundaries that had a high basal area. The average pixel value for bands 1-5 and 7 were calculated to yield the high basal area signature. Assuming that the pixels that are spectrally the furthest away from this signature represent low basal area stands, we calculated the Euclidian spectral distance from the high basal area signature to all other cells. The cells with the largest spectral distance were sampled and averaged to create the low basal area signature. Once again, the Euclidian spectral distance was calculated for all evergreen cells in the data set, but this time it was generated using the low basal area signature.

Separate regression models were generated and applied to the two scenes for which ground data were available by regressing stand basal area on the average Euclidian spectral distance for the stand. See table 1 for the regression model results.

We applied these models to the adjacent LTM scenes using their areas of overlap. Using the LTM-estimated basal area values, the high basal area regions were located on the adjacent, unprocessed scenes. Following the steps described above, the Euclidian spectral distances from the low basal area signature were calculated. Random sample points were generated in the unprocessed scene's area of overlap with the modeled scene.

Table 1.—*Basal area–spectral distance regression results*

LTM Scene	Adjusted R^2	Standard Error
Path 19, Row 37	0.82	19.20
Path 19, Row 38	0.80	21.11

Table 2.—*Adjacent basal area-spectral distance regression results*

LTM Scene	Parent Scene	Adjusted R^2	Standard Error
Path 19, Row 36	Path 19, Row 36	0.82	19.20
Path 18, Row 36	Path 19, Row 36	0.80	21.11
Path 18, Row 37	Path 19, Row 37	0.88	22.49
Path 18, Row 38	Path 19, Row 38	0.89	16.23
Path 17, Row 37	Path 18, Row 37	0.93	10.67
Path 17, Row 38	Path 18, Row 38	0.86	13.01

The LTM-estimated basal area from the modeled scene and the spectral distance from the unprocessed scene were recorded at each point. We combined samples into one-unit spectral distance classes where the basal area values in each sample in the class were averaged because of the large amount of noise in the LTM data set. The averaged sample data were then used to derive similar basal area—Euclidian spectral distance regression models. These steps were then repeated for all the unprocessed scenes adjacent to those LTM scenes. The results of the regression models applied to adjacent scenes are listed in table 2.

Image Segmentation

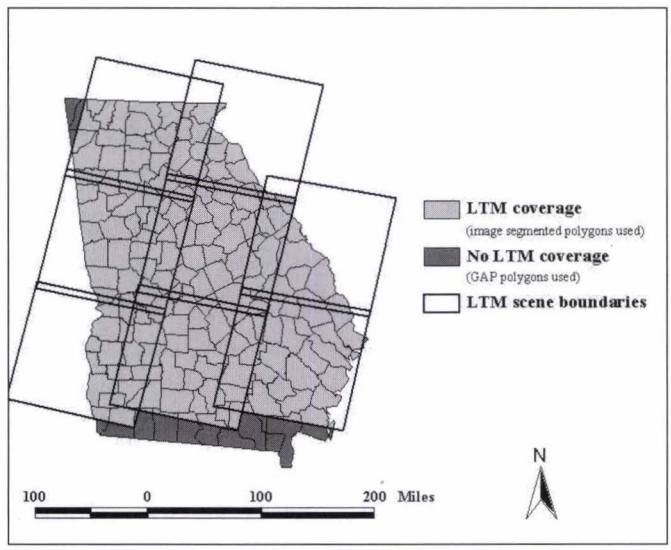
We used the GAP landcover data set as a first-stage sample of landcover type. We located on them the timbered and nontimbered areas, and ignored regions not classified as clearcut/sparse, deciduous forest, evergreen forest, mixed forest, or forested wetland. Then, we generated separate "evergreen," "deciduous," and "mixed" data sets by masking out all areas in the LTM scene that were not:

a. GAP-classified as *evergreen forest* for the evergreen LTM data set
b. GAP-classified as *deciduous forest* or *forested wetland* for the deciduous LTM data set
c. GAP-classified as *mixed forest* or *clearcut/sparse* for the mixed LTM data set.

As the second-stage sample of landcover type, the *evergreen*, *deciduous*, and *mixed* LTM data sets were then converted to homogenous polygons using an image segmentation module for ERDAS Imagine that was developed by the USDA Forest Service Remote Sensing Applications Center (Ruefenacht *et al.* 2002). This module iteratively compares the Euclidian spectral distance (Equation 1) between the first pixel in the image, the seed pixel, and the adjacent cells. If the Euclidian spectral distance between the two cells is less than or equal to the threshold value specified by the user, the pixel is assigned to the same region as the seed. The region is finished when there are no more cells adjacent to a member in the group that satisfy the threshold criteria. When the region is closed, the next seed pixel is selected and the process is repeated.

When vectorized, these image-segmented scenes produce a data set with over one million polygons, too many for ArcInfo to process in one data set. To reduce the data to a manageable size, the image-segmented LTM scenes were clipped using the county boundary. All processes described from this point on were applied on the county data sets. The image-segmented LTM data sets were then vectorized using ArcInfo's *GRIDPOLY* command. Finally, the *evergreen*, *deciduous*, and *mixed* LTM-generated polygon data sets were combined and cleaned

Figure 1.—*LTM scene boundaries.*



up by merging all polygons less than 10 acres with a larger adjacent polygon using the ArcInfo *ELIMINATE* command. GAP polygon data were substituted for those regions with no LTM coverage (fig. 1).

FIA Information Distribution

FIA information was distributed throughout the LTM-generated polygon data sets using two different methods. The first method iteratively assigned FIA information to the LTM-generated polygons using proximity to the FIA plot, current FIA plot type, and LTM-generated polygon type. The FIA data were distributed to the polygons until the sum of those polygons' acreage was within 99 percent of the FIA plot's per acre expansion factor or until there were no more polygons to fill. The data distribution criteria were applied to each FIA point before the next was applied. The criteria were as follows:

1) Assign FIA information to polygons that are within 1,000 meters of the plot location and of the same general type: evergreen, deciduous, or mixed.
2) Assign FIA information to polygons that are within 12,800 meters of the plot location and of the same general type: evergreen, deciduous, or mixed.
3) Assign information from evergreen FIA plots to evergreen or mixed polygons within 6,040 meters.
4) Assign information from deciduous FIA plots to deciduous or mixed polygons within 6,040 meters.
5) Assign information from mixed FIA plots to mixed or deciduous polygons within 6,040 meters.
6) Assign information from any FIA plot to polygons of any type within 6,040 meters.
7) Assign information from any FIA plot that has not been fully distributed to polygons of any type that have not been fully filled.

The second method iteratively assigned FIA information to the LTM-generated polygons using LTM-estimated pine basal area, polygon size, FIA volume per acre, and FIA peracre expansion factor. The evergreen polygons were first assigned the average (for the areas that fell within the stand) LTM-estimated basal area and ranked from highest to lowest; deciduous and mixed polygons were ranked according to acreage from highest to lowest, and the FIA plots were ranked from highest to lowest according to volume per acre. The FIA data were distributed to the polygons in proportions relative to polygon acreage and FIA plot per acre expansion factor. The data distribution criteria were as follows:

1) Assign information from evergreen FIA plots to evergreen LTM-generated polygons and then to mixed and hardwood polygons if more area is needed.
2) Assign information from deciduous FIA plots to deciduous LTM-generated polygons and then to any unassigned mixed and hardwood polygons if more area is needed.
3) Assign information from mixed FIA plots to mixed LTM-generated polygons and then to any unassigned deciduous or evergreen polygons.

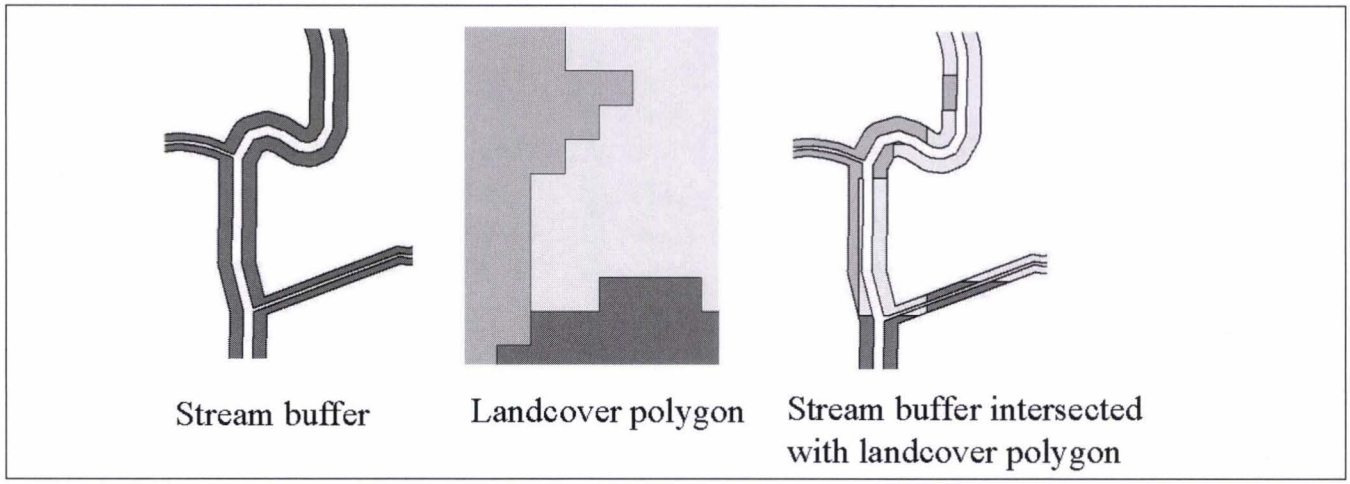
Buffer Creation

Following the criteria set forth in Georgia's BMPs, the hydrography, GAP, and NED data sets were used to generate riparian zone buffers that incorporated two buffer widths and three slope classifications. A total water mask was generated by first buffering the hydrology data set. Streams and rivers classified as perennial and those streams classified as trout streams in NARSAL's trout data set were buffered on each side by 15 feet, creating a 30-foot primary stream mask. The intermittent streams were buffered on each side by 12.5 feet, creating a 15-foot secondary stream mask. Those bodies of water contained in the GAP data set, all cells classified as "open water" or "coastal marsh," were vectorized and merged with the existing stream water masks to create the total water mask polygon data set. All areas within this mask were considered to be water.

Percent slope data sets were generated from the elevations contained in the NED data. Following the BMP guidelines for slope class, they were reclassified into three classes: 1) slight (< 20 percent), 2) moderate (20 – 40 percent), and 3) steep (> 40 percent). The slope data were then vectorized and incorporated into the water masks by intersection (fig. 2). This process produced a data set containing the polygons in the original water mask cut into smaller pieces where it and the slope data meet.

Transportation buffers were generated using similar methods. Cells classified as "transportation" in the GAP data set were buffered on both sides by 40 feet and 100 feet, excluding the "transportation" cells.

Figure 2.—*Stream buffer LTM polygon intersection.*



Data Intersection

The LTM-generated polygons were incorporated into the hydrology and transportation buffers by intersection (fig. 2). This process produced data sets similar to those created when the hydrology masks were intersected with the reclassified slope data. The final intersected data contain the buffer polygons cut into smaller pieces where they and the LTM-generated polygons meet. Attribute information from both data sets is stored for each data set, as well.

Results

Landsat Thematic Mapper satellite images and other spatial and tabular data were combined to assess the area, volume, age, and site quality of Georgia's riparian zone timberlands. LTM-derived Euclidian spectral distance–basal area models were calibrated using industrial ground data and applied statewide using LTM scene overlap. Homogenous evergreen, deciduous, and mixed forest polygons generated from the LTM images were populated with FIA tabular data using two different methods. The first method was based on the proximity to the FIA plot. Polygons, of the same type as the FIA plot, were assigned data, starting with the closest and proceeding away until the FIA per acre expansion factor-based criteria were met. The other method assigned FIA data to the polygons using ranking criteria. Evergreen polygons were ranked according to LTM-estimated basal area, hardwood and mixed forest polygons

Table 3.—*Intersection data set descriptions*

Data Set Name	Description
Primary hydro/proximity	Primary hydrology buffer intersected with landcover containing FIA information assigned using the proximity criteria
Primary hydro/rank	Primary hydrology buffer intersected with landcover containing FIA information assigned using the ranking criteria
Secondary hydro/proximity	Secondary hydrology buffer intersected with landcover containing FIA information assigned using the proximity criteria
Secondary hydro/rank	Secondary hydrology buffer intersected with landcover containing FIA information assigned using the ranking criteria
100-foot road/proximity	100-foot road buffers intersected with landcover containing FIA information assigned using the proximity criteria
100-foot road/rank	100-foot road buffers intersected with landcover containing FIA information assigned using the ranking criteria
40-foot road/proximity	40-foot road buffers intersected with landcover containing FIA information assigned using the proximity criteria
40-foot road/rank	40-foot road buffers intersected with landcover containing FIA information assigned using the ranking criteria

according to acreage, and FIA data by plot volume per acre measures. The highest ranked polygons were assigned information from the highest ranked FIA plots (of the same type), and so on. Riparian zone buffers were generated using two different buffer distances incorporating three different slope classes. Similar transportation buffers were generated, buffering GAP-classified roads by 40 feet and 100 feet. LTM-generated polygons and riparian and road buffers were incorporated by intersection (fig. 2). Eight intersection data sets were generated (table 3), each containing the buffer polygons split where they and the landcover polygons meet.

Literature Cited

Georgia Forestry Commission. 1999. Georgia's best management practices for forestry. Georgia Forestry Commission.

Ruefenacht, B.; Vanderzanden, D.; Morrison, M.; Golden, M. 2002. New technique for segmenting images. RSAC Program Manual. U.S. Department of Agriculture, Forest Service, Remote Sensing Application Center. 14 p.

Assessing the Effects of Forest Fragmentation Using Satellite Imagery and Forest Inventory Data

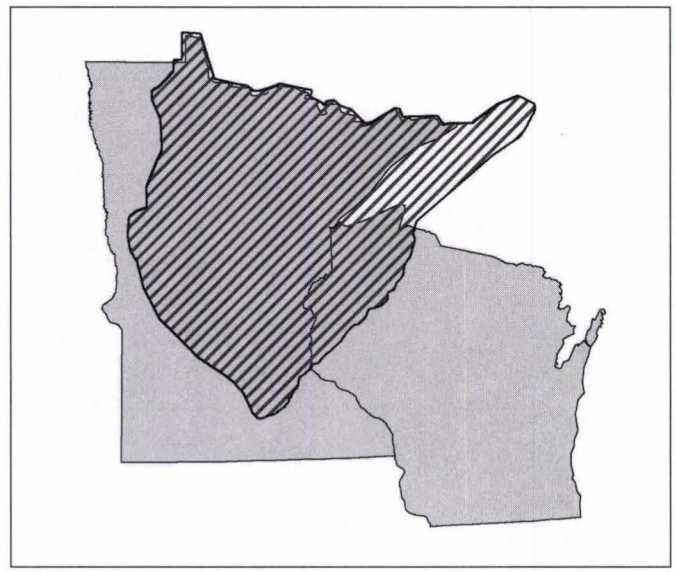
Ronald E. McRoberts[1] and Greg C. Liknes[1]

Abstract.—For a study area in the North Central region of the USA, maps of predicted proportion forest area were created using Landsat Thematic Mapper imagery, forest inventory plot data, and a logistic regression model. The maps were used to estimate quantitative indices of forest fragmentation. Correlations between the values of the indices and forest attributes observed on forest inventory plots were estimated. One interesting result was a statistically significant negative correlation between total forest area and number of tree species per unit forest area.

Montreal Process

Over the last decade, natural resource managers, the scientific community, and the general public have voiced serious concerns regarding the status of and emerging trends in the world's forests. In 1993, the government of Canada began a series of meetings to develop scientifically rigorous methods for evaluating forest management. These meetings led to the Montreal Process criteria and indicators for environmental and ecological assessments of forest sustainability. One criterion describes conditions or processes by which sustainable forest management may be evaluated and is further characterized by a set of indicators that are monitored periodically to assess change. Four of the seven Montreal Process criteria deal with forest conditions and attributes: (1) conservation of biological diversity, (2) maintenance of productive capacity of forest ecosystems, (3) maintenance of forest ecosystem health and vitality, and (4) maintenance of forest contribution to global carbon cycles. Forest fragmentation affects several of these criteria. Our study sought to evaluate the effects of forest fragmentation on a variety of forest stand attributes. The study was conducted in the North Central region of the United States and included portions of Minnesota, Wisconsin, and Michigan (fig. 1).

Figure 1.—*Study area.*



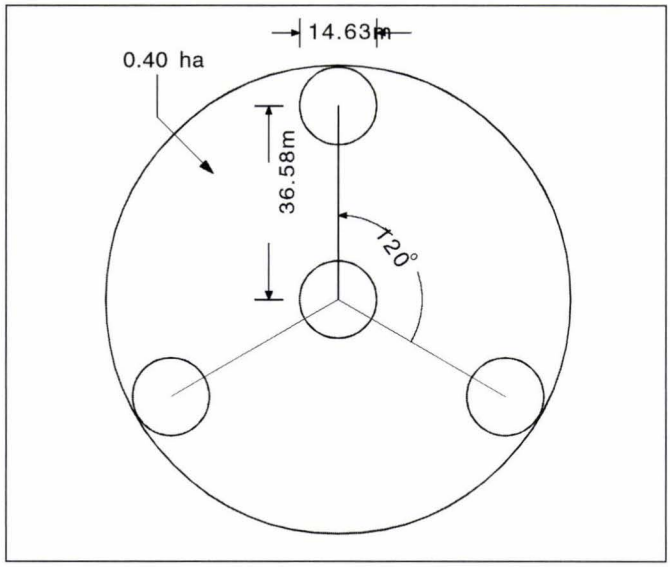
Data and Methods

Inventory Plot Data

The Forest Inventory and Analysis (FIA) program of the Forest Service, U.S. Department of Agriculture, has established an array of permanent field plots using a systematic sampling design. In the North Central region, a fixed proportion of plots are measured each Federal fiscal year (01 October to 30 September). Plots measured in the same Federal fiscal year comprise a single panel of plots, and panels are measured on a rotating basis. In aggregate, over a complete measurement cycle of 5 years, a plot represents approximately 2,403 ha. In general, locations of forested or previously forested plots are determined using global positioning system receivers, while locations of nonforested plots are determined using digitization methods. Each field plot consists of four 7.31-m radius circular

[1] Mathematical Statistician and Computer Specialist, respectively, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

Figure 2.—*Forest Inventory and Analysis standard plot design.*



subplots configured as a central subplot and three peripheral subplots with centers located 36.58 m and azimuths of 0 degrees, 120 degrees, and 240 degrees from the center of the central subplot (fig. 2). For each tree, field crews report species, live or dead status, and diameter at breast height (d.b.h.) (1.37 m). Regression models are used with observed d.b.h. as an independent variable to predict the volumes and biomass for individual trees. In addition, field crews note evidence of both natural and human disturbances, estimate the number of seedlings, and estimate the proportions of each subplot that satisfy specific land use conditions. Subplot estimates of forest land proportions are obtained by aggregating these land use conditions into forest and nonforest uses. Plot estimates of number of species, number of live and dead trees, biomass in live and dead trees, live tree volume, number of seedlings per unit area, and average stand diameter were obtained by aggregating individual trees and subplots. Observations for 1,185 plots were available for the 1999, 2000, and 2001 panels.

Satellite Imagery Classification

Landsat TM imagery, classified according to forest and nonforest, was used to quantify fragmentation. The images included data for three dates, consisted of 30 m x 30 m pixels for bands 1-5 and band 7, and were geo-referenced to Albers Equal Area projection, NAD 83.

The first step was to calibrate a model for predicting the proportion of forest land for each image pixel. Because forest land proportions are always in the closed interval [0,1], it is appropriate to select a model with mathematical properties that restrict predictions to the same interval. The logistic model is often used with such data and was selected for this study to describe the relationship between observed forest land proportion for FIA subplots and the spectral values of the pixels containing the subplot centers,

$$E(Y_k) = \frac{1}{1+\exp\left(\beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \ldots + \beta_p X_{pk}\right)}$$

where E(.) denotes statistical expectation, Y_k is the forest land proportion for the k^{th} subplot, X_{jk} is the spectral value for the j^{th} band for the pixel containing the center of the k^{th} subplot, and the βs are parameters to be estimated. For each study area, all possible band combinations were compared according to root mean square error, and the combination with the smallest root mean square error was selected.

After calibration, the models were used to predict forest land proportion for each pixel in the study area. In accordance with the practice of other mapping agencies, pixels with proportion forest land predictions less than 0.25 were designated nonforest, and pixels with forest land predictions equal to or greater than 0.25 were designated forest. Slightly less than 90 percent of the nonforest plots were correctly classified, and slightly more than 90 percent of the forest plots were correctly classified.

Correlation and Validation Analyses

For the 0.4-ha circle circumscribing the four subplots of each FIA plot, four measures of forest fragmentation were taken: forest edge length, edge forest area, interior forest area, and total forest area. Forest edge length was calculated as the total length within the 0.4-ha circle of the forest/nonforest boundary between pixels classified as forest and pixels classified as nonforest; edge forest was calculated as the total area within the circle of forest pixels within two pixel widths of the forest/nonforest boundary; interior forest area was calculated as the area within the circle of forest pixels greater than two pixel widths from the forest/nonforest boundary; and total forest area was calculated as the sum of edge and interior forest area. The three

Table 1.—*Estimated correlations between measures of forest fragmentation and forest stand attributes*[1]

Forest attribute	Fragmentation index		
	Edge forest area	Total forest area	Interior forest area
No. species	0.03	**-0.19**	**-0.44**
No. live trees	**-0.19**	**0.14**	-0.05
Live biomass	-0.05	-0.03	**-0.20**
Seedlings	-0.01	-0.03	**-0.10**
Mean d.b.h.	**0.11**	**-0.13**	**-0.10**

[1] Estimates in bold are statistically significant at α=0.01.

area measures were divided by the total area of the 0.4-ha circle and were analyzed as proportions of that area.

The first stage of analysis consisted of simple correlation analyses between the estimates of forest attributes and the four measures of forest fragmentation (table 1). Plots with no forest land within the 0.4-ha circle were excluded from the analyses as were plots with evidence of human-caused disturbance, leaving 1,185 plots. The forest attribute measures previously described were all divided by the total forest area to scale estimates to a per unit forest area basis. The high negative correlations between number of species per unit forest area and total forest area were of particular interest and suggested that the number of species per unit of forest area may be greater when forest fragmentation is greater. This result warrants further investigation because lesser fragmentation and greater species richness, defined as the number of species per unit area, are both generally viewed as positively affecting forest sustainability. The observed result, however, suggests that greater species richness is associated with greater, not lesser, fragmentation.

The second stage of analysis focused on validating the inverse relationship between number of tree species per unit of forest area on FIA plots and the proportion of the 0.4-ha circular plot that was forested (fig. 3). The primary issue was determining if the large negative correlation could be due to artifacts resulting from expressing number of species found on a plot on a per unit forest area basis. First, if even small forested areas are saturated with species, then the decrease in number of species per unit of forest area as the proportion of plot forest area increases could be attributed to dividing a relatively con-

Figure 3.—*Number of species per unit area on FIA plots versus proportion of plot in forest area.*

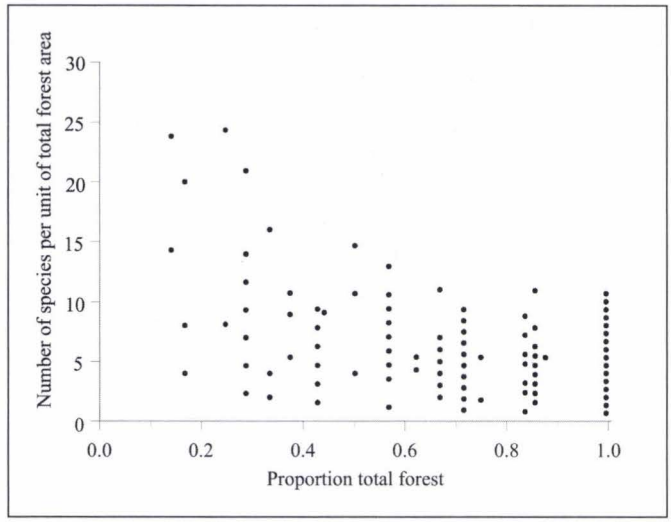
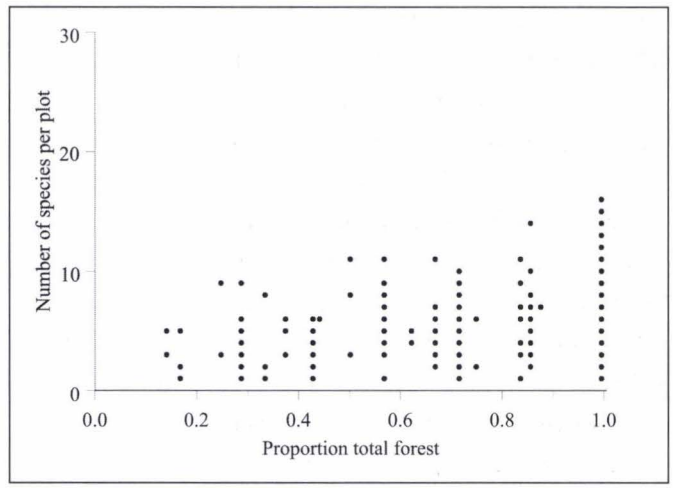


Figure 4.—*Absolute number of tree species found on FIA plots versus proportion of plot in forest area.*



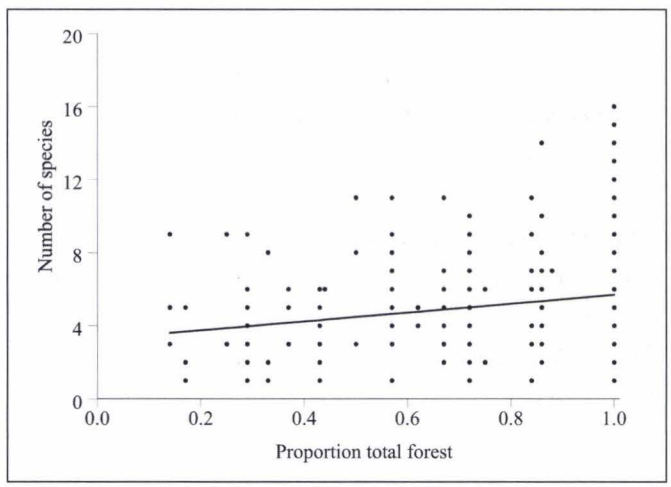


Table 2.—*Number of different species*

Proportion forest	No. plots	Cumulative no. species
0.00-0.09	13	25
0.10-0.29	22	31
0.30-0.49	21	27
0.50-0.69	31	35
0.70-0.89	100	39
0.90-1.00	998	67

stant number by an increasing number. A graph of the data, however, reveals that the absolute number of species increases as the proportion of forest area increases (fig. 4). Second, if only the same small number of species ever occur on plots with small proportion forest areas, then the large number of species per unit forest area could be attributed to dividing this same small absolute number of species by small proportion forest areas. However, the absolute numbers of species found on all plots listed by categories of proportion forest area appear to dispel this possibility (table 2). In addition, these latter results confirm that absolute numbers of species per plot increases as the proportion of forested area on the plot increases. Over all plots with proportion forest area between 0.00 and 0.0, 25 species were found, while on plots with proportion forest area between 0.10 and 0.29, 31 species were found. These numbers are similar to those indicated by figure 4 for plots with small proportion forest areas. Although there are additional sampling issues associated with estimating number of species per unit area that have not been addressed in this study, it appears that the large, statistically significant, negative correlation between tree species per unit forest area and total forest area may be a genuine phenomenon and not simply an arithmetic artifact.

Summary

In brief summary, a forest/nonforest classification of satellite imagery provided a good means of measuring forest fragmentation such as edge length, edge forest area, and interior forest area. Further, these measures may be easily correlated with estimates of forest attributes such as tree species per unit area obtained from forest inventory plot data. The most significant result was the negative correlation between number of tree species per unit forest area, a measure of species richness, and proportion forest area, a measure of forest fragmentation. This correlation suggests that two positive indicators of forest sustainability, less forest fragmentation and greater tree species richness, perhaps should not be expected to occur simultaneously.

Comparing Forest/Nonforest Classifications of Landsat TM Imagery for Stratifying FIA Estimates of Forest Land Area

Mark D. Nelson[1], Ronald E. McRoberts[2], Greg C. Liknes[1], and Geoffrey R. Holden[1]

Abstract.—Landsat Thematic Mapper (TM) satellite imagery and Forest Inventory and Analysis (FIA) plot data were used to construct forest/nonforest maps of Mapping Zone 41, National Land Cover Dataset 2000 (NLCD 2000). Stratification approaches resulting from Maximum Likelihood, Fuzzy Convolution, Logistic Regression, and k-Nearest Neighbors classification/prediction methods were superior to an unstratified, simple random sampling approach for producing stratum weights used to lower the variance of estimates of FIA mean proportion forest land. The stratification approaches were comparable to one another.

Each of the Forest Inventory and Analysis (FIA) units of the U.S. Department of Agriculture is required to report estimates of forest land area for their respective regions every 5 years. These estimates are obtained by multiplying total area inventoried by the mean proportion forest land estimated from forest inventory field plots. Forest land, as defined by FIA, includes commercial timberland; some pastured land with trees; forest plantations; unproductive forest land; and reserved, noncommercial forest land. Additional criteria for FIA forest land include 10 percent minimum stocking (5 percent canopy cover for several western woodland types where stocking cannot be determined), minimum area of 0.405 ha (1 acre), and minimum continuous canopy width of 36.58 m (120 ft) (USDA 2002). National FIA precision standards limit the allowable error for estimates of forest land area. Due to natural variability among plots and budgetary constraints, sample sizes sufficient to satisfy national FIA precision standards are seldom achieved. To meet these standards, a stratified estimation approach is used to reduce errors of estimates.

Traditionally, FIA has interpreted a set of aerial photo plots to obtain stratum weights (Phase 1). A subset of Phase 1 plots was measured in the field (Phase 2). This double sampling approach produced estimates that attained national precision standards for forest area (Hansen 1990). However, stratification based on aerial photography has some limitations: It is labor intensive and subjective; photos are expensive and cumbersome to transfer, handle and store, the interpretation is prone to bias when field plots are interpreted differently than nonfield plots; and the photos can be of variable quality and timeliness (McRoberts *et al.* 2002a).

To overcome these limitations, FIA is developing methods of satellite image classification for creating Phase 1 strata. Image pixels within an area of interest are divided into homogeneous classes, based on predictions of land cover. These classes form strata for stratified estimation of Phase 2 data. Stratified estimation can yield increases in precision, even when within-stratum sampling intensities are independent of stratification (McRoberts *et al.* 2002a). Advantages of using satellite imagery for stratification include the following: the resulting coverage is "border-to-border," not a sample of the analysis area; stratum weights are obtained easily from pixel counts; Phase 2 plots are assigned objectively to strata using a geographic information system (GIS); and satellite image stratification can be much cheaper and faster than photo-based stratification. The question is, How precise are estimates based on these stratifications—do they satisfy allowable error standards?

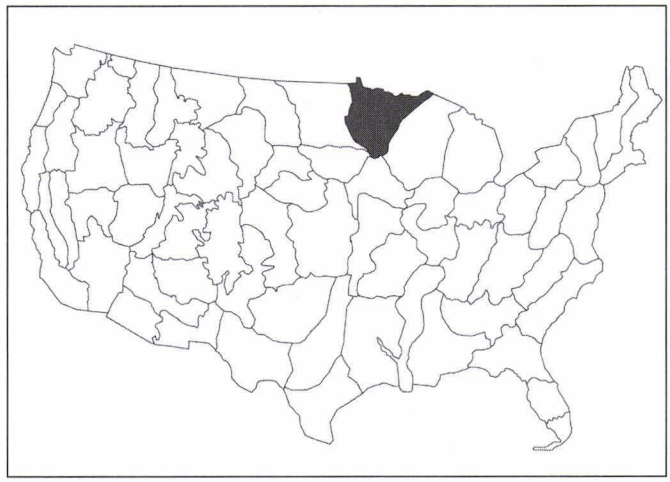
The North Central Research Station (NCRS) FIA program (NC-FIA) measures plots every 5 years across 11 States in the upper Midwest and Great Plains. A stratification based on Landsat-5 Thematic Mapper (TM) or Landsat-7 Enhanced Thematic Mapper Plus (ETM+) imagery will require processing of approximately 125 scenes in the NC-FIA region. Thus, a need exists for rapid processing of TM imagery for creation of Phase 1 strata used in stratified estimation.

[1] GIS/RS Specialist, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.
[2] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

The Multi-Resolution Land Characterization (MRLC) consortium of the U.S. Geological Survey has mosaicked Landsat TM and ETM+ imagery into regional mapping zones. These National Land Cover Dataset 2000 (NLCD 2000) mapping zone image data allow for more efficient image classification than when individual TM scenes are used (Homer and Gallant 2001).

The objective of our study was to compare stratifications produced from classifications of an NLCD 2000 mapping zone data set using four approaches: (1) maximum likelihood supervised classification, (2) maximum likelihood fuzzy convolution classification, (3) a classification using a logistic regression modeling approach, and (4) a classification using a non-parametric, k-Nearest Neighbors (k-NN) approach.

Figure 1.—*NLCD 2000 Mapping Zones and the Zone 41 study area (gray).*



Study Area

The study was conducted within NLCD 2000 Mapping Zone 41, hereafter referred to as Zone 41. This zone encompasses 181,000 square kilometers in portions of eastern Minnesota, northwestern Wisconsin, and northwestern Michigan (fig. 1). The area is characterized by prairie agriculture, a diverse mixture of forest land including both coniferous and deciduous species, and a portion of Lake Superior.

Data

Satellite Imagery

Satellite data for Zone 41 are from TM and ETM+ images (fig. 2). This set of images has the following attributes: (1) 30 m x 30 m pixels from bands 1-5 and band 7; (2) absolute radiance units scaled to 8 bits; (3) processing to level 10: radiometrically corrected, using satellite model and platform/ephemeris information, rectified using ground control points and digital elevation model terrain correction, and resampled, using cubic convolution with resulting root mean square error less than 8.5 m; and (4) geo-referencing to USGS Albers Equal Area projection, NAD83. Image data include optical band values and tasseled cap transformations for three seasons: spring, leaf-on (summer) and leaf-off (late fall / early winter). Kauth and Thomas (1976) introduced the "tasseled cap" transformation of Landsat Multispectral Scanner (MSS) imagery as an easily visualized, three-dimensional construct of the most important phenomena of crop development. Key forest attributes, e.g., species, age, and structure also may be revealed by the transformation (Cohen *et al.* 1995). Crist and Cicone (1984) modified the tasseled cap transformation for TM imagery. Images resulting from the transformation collectively explain about 97 percent of spectral variance within a scene while reducing six original TM bands to three components: brightness, greenness, and wetness.

Figure 2.—*NLCD 2000 Mapping Zone 41, leaf-on, true color image; TM/ETM+ bands 1 (blue), 2 (green), and 3 (red).*

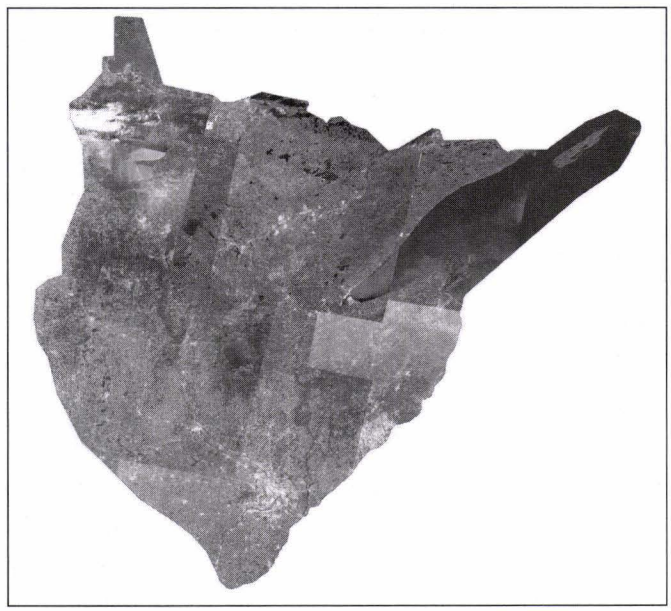
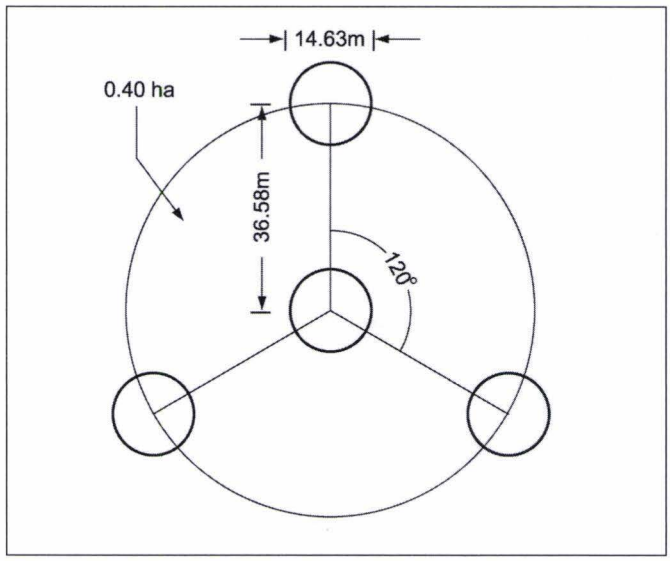


Figure 3.—*FIA Phase 2 plot design.*



FIA Plot Data

Under the FIA program's annual inventory system, field plots are established in permanent locations using a systematic sampling design with each plot representing 2,403 ha (McRoberts 1999). Approximately 20 percent of the plots in each State are measured annually. Locations of forested or previously forested plots are captured using global positioning system (GPS) receivers. Locations of nonforested plots are determined using digitization methods.

Each field plot consists of four 7.31-m (24-ft)-radius circular subplots, configured as a central subplot and three peripheral subplots with centers separated by 36.58 m (120 ft) at azimuths of 0°, 120°, and 240° from the center of the central subplot (fig. 3).

Observations obtained by field crews include the proportions of subplot areas that satisfy specific land use conditions. Plot-wise proportions of forest and nonforest land are determined by computing the mean proportions of these two land uses across the four subplots. Measurements from 5,939 plots associated with cloud-free areas of Zone 41 satellite imagery were used in this study: 5,242 from Minnesota (years 1999–2001) and 697 from Wisconsin (years 2000–2001). Of the measured plots, 2,439 were completely forested, 94 were partially forested, and 3,406 were nonforested.

Methods

Mapping

Seven stratification maps were produced using variations of two classification methods and two prediction methods: (1) maximum likelihood (ML), (2) fuzzy convolution (Fuzz), (3) logistic regression modeling (Log) and (4) k-Nearest Neighbors (k-NN), respectively. Names of stratification approaches incorporate notation for the classification or prediction method (e.g., ML), the number of input training classes for the two classification methods, and the presence or absence of edge strata (table 1).

ML

ML classifications were produced using training data from the following tasseled cap images: spring brightness, spring greenness, spring wetness, leaf-on greenness, and leaf-off brightness.

Table 1.—*Approaches for producing stratified estimates of mean proportion forest land, NLCD 2000 Mapping Zone 41; Nonforest (NF), Nonforest Edge (NFE), Forest (F), Forest Edge (FE), Terrestrial Nonforest (TNF), and Water (W) strata*

Stratification approach	Classification/ prediction method	Inputs	Strata					
			NF	NFE	F	FE	TNF	W
ML2Edge	ML	Nonforest, forest	X	X	X	X		
ML3	ML	Nonforest, forest, water			X		X	X
ML3Edge	ML	Nonforest, forest, water	X	X	X	X		
Fuzz3	Fuzz	ML3, distance			X		X	X
Fuzz3Edge	Fuzz	ML3, distance	X	X	X	X		
LogEdge	Log	Proportion forest land use	X	X	X	X		
k-NNEdge	k-NN	Proportion forest land use	X	X	X	X		

These image layers were selected to correspond with those identified as the "best" bands for k-NN analysis (see below). Chen and Stow (2002) recommend using single pixels for training because pixels that are contiguous or close together may exhibit spatial autocorrelation. If training data are collected from autocorrelated pixels, the variance of this training data tends to be reduced. This may produce biased training signatures that are less representative. Therefore, we used single pixels associated only with central subplots, which are spatially separated from pixels associated with central subplots of other plots.

Based on proportion forest land use, each subplot was categorized as nonforest (< 0.25) or forest (≥ 0.25) before performing the image classifications. The 0.25 minimum threshold for proportion forest land is comparable to the definition of forest land currently used for the Natural Resources Conservation Service (NRCS) Natural Resources Inventory (NRI) (Lessard *et al.* 2003) and is approximately equivalent to FIA's requirement of 10 percent minimum stocking. In comparison, the NLCD definition of forest is land that has 20 percent or more forest cover (tree crown cover or crown closure); Anderson *et al.* (1976) define forest land as having 10 percent or more tree-crown density (crown closure percentage).

Nonforest and forest class signature files were created by appending individual spectral signatures from image pixels associated with each plot location. Due to the cumbersome nature and long processing time associated with nearly 6,000 individual signatures (1 pixel for each central subplot), a guided clustering technique was used (Bauer *et al.* 1994, Lillesand *et al.* 1998). Using this approach, two ISODATA unsupervised classifications were performed, one for pixels associated with central subplots defined as nonforest and one for pixels associated with central forested subplots. Parameters for both ISODATA classifications were as follows: classes = 5, iterations = 20, convergence threshold = 0.98. The resulting five signatures each for nonforest and forest were subsequently merged into one signature for each of the two classes. A classification based on these two signatures was used to produce the ML2Edge stratification.

Merging the five nonforest signatures into a single signature resulted in a bimodal distribution of tasseled cap data—a violation of the requirement for normal data distribution when performing a maximum likelihood classification. Therefore, ISODATA classes 1 and 2 and ISODATA classes 3, 4, and 5 were merged into two normally distributed signatures, characteristic of water and terrestrial nonforest, respectively. Water, terrestrial nonforest, and forest signatures were used to complete a supervised classification using the ML parametric rule and a fuzzy classification option. Output consisted of the three best classes per pixel with a corresponding distance image. Layer one of the fuzzy classification output represents the most likely class for each pixel and was used to produce the ML3 stratification. Following classification, water and terrestrial nonforest pixels were recoded as a single nonforest class, resulting in a classification used for the ML3Edge stratification. ML3Edge is comparable to ML2Edge, but the classification used for ML3Edge conforms to the requirement for using normally distributed data in ML analyses.

Fuzz

Fuzzy convolution is a technique that creates a classification layer by "...calculating the total weighted inverse distance of classes in a window of pixels and assigning the center pixel the class with the largest total inverse distance summed over the entire set of fuzzy classification layers" (Pouncey *et al.* 1999). Whereas classes with higher distance values may change to a neighboring value, classes with a very small distance value will remain unchanged. The result is a context-based classification with reduced speckle. The Fuzz3 stratification was produced using the ML3 fuzzy classification and distance layers described above. The distance neighborhood weighting was calculated within a 3-by-3 window with the central pixel weighted by 1.0, four vertical/horizontal pixels weighted by 0.646, and four diagonal pixels weighted by 0.500. Following classification, water and terrestrial nonforest classes were merged into a single nonforest class. This nonforest class, along with the Fuzz3 forest class was used to produce the Fuzz3Edge stratification.

Log

A logistic regression model with mathematical properties that restrict predictions in the interval [0,1] was selected to accommodate forest land proportions, which also are constrained to the interval [0,1]. McRoberts and Liknes (2002) describe this approach for estimating proportion forest land. In this approach, all four subplots associated with each plot were used.

In brief, a three-step process was used to select spectral bands for inclusion in models. First, the data were transformed to permit use of a linear model, which accelerated the computer processing speed for selecting optimal image band combinations. Second, simple linear regression models were fit to the transformed observations. Third, Logistic models using the five best combinations of bands with smallest Root Mean-Square Error (RMSE) were fit to the forest land proportion observations using weighted nonlinear regression where the weights reflected the correlations among observations of subplots within the same plot. The model using the band combination and the corresponding parameter estimates obtained from the nonlinear analyses was used to create a map of forest land proportion predictions by calculating a prediction for each pixel. Each of the five best Log models contained three spectral bands. The best combination of bands identified for the Log approach were leaf-off near infrared (TM band 4), leaf-off normalized difference vegetation index (NDVI), and leaf-on NDVI. Continuous estimates of proportion forest cover were divided into forest and nonforest land cover strata using the same definitions as for ML, described above.

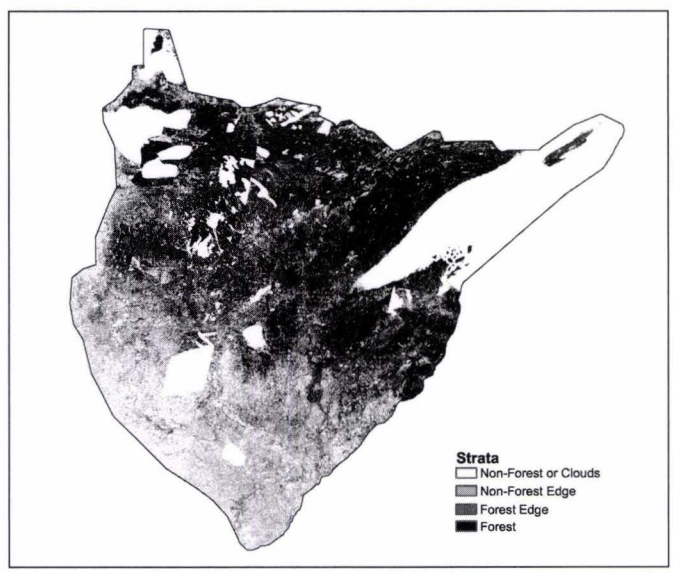
k-NN

The k-Nearest Neighbors technique is a nonparametric approach for predicting values of point variables. Similarity is based on a covariate space between the point and other points with observed values of the variable. McRoberts *et al.* (2002b) describe the k-NN methodology used in this study to create continuous estimates of forest cover. The observed values are the forest cover proportions for each FIA subplot. The forest cover prediction for each pixel is based on the average proportion forest cover of the k subplots with corresponding pixel spectral values nearest to that of the pixel in question. Unweighted Euclidean distance was used to identify those k-neighbors nearest in spectral space. The value of k was based on the number that minimized RMSE for each combination of spectral bands. The leaving-one-out method was used to obtain RMSE of forest land proportion. The five combinations of spectral bands with smallest resulting RMSE were used to predict proportion forest cover for each image pixel. The five best k-NN calibrations had three to five bands. The best calibration contained five bands (tasseled cap: spring brightness, spring greenness, spring wetness, leaf-on greenness, leaf-off brightness) and had a value of k=24. Continuous estimates of proportion forest cover were divided into forest and nonforest land cover strata using the same definitions as for ML and Log, described above.

Classifications based on ML, Fuzz, Log, and k-NN methods were processed further using clump and eliminate functions (Pouncey *et al.* 1999) to remove isolated single pixels and groups of pixels of one class when their contiguous area was smaller than < 0.405 ha (FIA definition of 1-acre minimum area).

Hansen and Wendt (2000) and McRoberts *et al.* (2002a) reported that the efficiency of stratifications was improved when separating edge strata from forest and nonforest strata at forest/nonforest boundaries. Therefore, before performing stratified estimation, image pixels were processed to subdivide both forest and nonforest classes into interior and edge classes. Pixels of either forest or nonforest class that are within a 2-pixel distance (60 m horizontal/vertical distance, 85 m diagonal distance) from a forest/nonforest boundary are labeled as edge pixels. All other pixels are considered non-edge and retain their original designation as forest or nonforest. This procedure resulted in the following classes representing four strata: nonforest, nonforest edge, forest, and forest edge (fig. 4). Edge pixels were not identified for ML3 or Fuzz3 stratification approaches (table 1).

Figure 4.—*ML3Edge stratification: nonforest, nonforest edge, forest, and forest edge strata.*



Stratified Estimation

Stratified estimates of mean plot forest land proportion, $\overline{P}$, and estimated variance, $\hat{Var}(\overline{P})$,, are calculated using formulae for stratified analysis (Cochran 1977):

$$\overline{P} = \sum_{h=1}^{L} w_h \overline{P}_h \quad (1)$$

and

$$\hat{Var}(\overline{P}) = \sum_{h=1}^{L} w_h^2 \frac{\hat{\sigma}_h^2}{n_h} \quad (2)$$

where h = 1,...,L denotes stratum; w_h is the hth stratum weight; $\overline{P}_h$ is the mean forest land proportion for plots assigned to the hth stratum; n_h is the number of plots assigned to the hth stratum; and $\hat{\sigma}_h^2$ is the within-stratum variance for the hth stratum. Variance estimates obtained using (2) ignore the slight effects due to finite population correction factors and to variable, rather than fixed, numbers of plots per stratum.

Stratum weights were determined as the proportions of pixels assigned to strata. Each FIA plot was assigned to one stratum. We avoided the mathematical complexity associated with spatial correlation among four subplots by assigning plots, rather than subplots, to strata. For this study, only the pixels associated with central subplot locations (plot centers) were used for assigning strata to plots.

Comparisons

Estimates of mean forest land proportion and the standard error of the mean were calculated assuming simple random sampling (SRS) for comparison purposes. Stratified analyses were conducted using either three or four strata, as defined in table 1. For the Log and k-NN analyses, stratifications from only the single best models (based on the smallest standard error of mean proportion forest land) were used (McRoberts 2002).

Results

Zone 41 estimates of mean proportion forest land were similar among all stratified approaches and were slightly smaller than the SRS estimate. Standard errors (SEs) of these estimates were noticeably smaller for the stratified approaches than for the SRS unstratified approach. For ML classifications, replacing bimodally distributed spectral signatures with signatures of normal distribution did not change estimates or standard errors of estimates. Standard errors based on stratifications with four strata were indistinguishable for ML, Fuzz, and Log approaches, and were slightly larger for the k-NN approach. Standard errors were slightly smaller when using stratifications with four strata (nonforest, nonforest edge, forest, and forest edge) than when using stratifications with three strata (water, terrestrial nonforest, and forest) for both ML and Fuzz approaches (table 2).

Discussion

Zone 41 stratifications derived from image classifications are useful for reducing standard errors of mean proportion forest land estimates. None of the stratification approaches is superior to the others, but all are superior to the unstratified SRS

Table 2.—*Simple random sampling and stratified estimation of mean proportion forest land, NLCD 2000 Mapping Zone 41*

Estimate	Stratification approach	Mean proportion forest land	Standard error
Unstratified	SRS	0.41	0.0061
Stratified	ML2Edge	.37	.0038
Stratified	ML3	.38	.0040
Stratified	ML3Edge	.38	.0038
Stratified	Fuzz3	.38	.0039
Stratified	Fuzz3Edge	.38	.0038
Stratified	LogEdge	.38	.0038
Stratified	k-NNEdge	.38	.0039

approach. A vegetation index (NDVI) and a tasseled cap transformation were more useful than TM/ETM+ optical band data for the k-NN and Log approaches, respectively.

In two related studies, both within Zone 41, standard errors of estimates from the Log approach were smaller than for ML and larger than for k-NN approaches in a less heavily forested area in central Minnesota (Nelson *et al.* 2002) but smaller than for the k-NN approach in a more heavily forested area in northeastern Minnesota (McRoberts 2002).

The ML, Fuzz, Log, and k-NN approaches all require acquisition and processing of satellite imagery. A visual comparison revealed the Fuzz approach produced smoothed variants of ML classifications, as expected. Although ML and Fuzz approaches are available as standard components of image processing software, Log and k-NN approaches are less accessible. A tool currently being developed to allow k-NN processing directly from ERDAS Imagine software will allow more widespread use of k-NN for processing satellite imagery. Although the k-NN technique is conceptually easy to implement, careful attention must be paid to its calibration to achieve optimal results. In addition, several precautions should be observed when using the k-NN technique (McRoberts *et al.* 2002b).

More work is needed to determine the optimal threshold for producing stratifications from continuous estimates of proportion forest (e.g., Log and k-NN estimates). Rather than producing a binary stratification (nonforest vs. forest) a stratification with multiple strata could be tested, e.g., 0.0 – 0.2, 0.2 – 0.4, 0.4 – 0.6, 0.6 – 0.8, and 0.8 – 1.0 proportion forest land. Since the estimate of proportion forest land follows a continuum, could we stratify along a comparable continuum? Multiple iterations of stratified estimation could be run, selecting those thresholds where SE's are minimized. If FIA policy requires a nonforest/forest stratification, the above methods could provide a benchmark of potential SEs to gauge performance of nonforest/forest stratification methods.

Zone 41 Landsat TM and ETM+ imagery consists of three seasonal mosaics of adjacent, semi-overlapping scenes from 1999-2001. Spring, leaf-on, and leaf-off imagery include scenes from early March through early May, early June through early August, and mid October through mid November, respectively. When producing zonal mosaics, MRLC gave precedence to selecting overlapping portions of scenes to those dates with least cloud cover. Despite these and other image processing steps employed by MRLC, some cloud cover and scene-related radiometric variability is evident within each seasonal mosaic (fig. 2). The classification/estimation of any portion of an NLCD 2000 mapping zone (e.g., individual TM/ETM+ scene) depends upon the selection of plots and their associated pixels. Future study could compare classifications of individual scenes using only the pixels associated with plots in that scene with their corresponding areas subset from classifications of zonal mosaics using pixels associated with plots distributed throughout the zone. When conducting stratified analyses requiring complete coverage of an area (assigning every pixel to a stratum for determining stratum weights), additional image processing of zonal mosaics may be required to eliminate cloud cover.

Acknowledgment

The authors wish to thank Dale Gormanson, Andrew Lister, and Kathy Ward for reviewing this manuscript.

Literature Cited

Anderson, J.R.; Hardy, E.E.; Roach, J.T.; Witmer, R.E. 1976. A land use and land cover classification system for use with remote sensor data. Washington, DC: U.S. Geological Survey. Prof. Pap. 964. (Revision of the land use classification system as presented in U.S. Geological Circular 671).

Bauer, M.E.; Burk, T.E.; Ek, A.R.; *et al.* 1994. Satellite inventory of Minnesota forest resources. Photogrammetric Engineering & Remote Sensing. 60(3): 287–298.

Chen, D.M; Stow, D. 2002. The effect of training strategies on supervised classification at different spatial resolutions. Photogrammetric Engineering & Remote Sensing. 68(11): 1155–1161.

Cochran, W. 1977. Sampling techniques. 3d ed. New York, NY: John Wiley and Sons. 428 p.

Cohen, W.B.; Spies, T.A.; Fiorella, M. 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. International Journal of Remote Sensing. 16: 721–746.

Crist, E.P.; Cicone, R.C. 1984. Application of the tasseled cap concept to simulated Thematic Mapper data. Photogrammetric Engineering & Remote Sensing. 50: 343–352.

Hansen, M. 1990. A comprehensive sampling system for forest inventory based on an individual tree growth model. St. Paul, MN: University of Minnesota. 268 p. Ph.D. dissertation.

Hansen, M.; Wendt, D. 2000. Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In: McRoberts, R.; Reams, G.; Van Deusen, P., eds. Proceedings, 1st annual forest inventory and analysis symposium; 1999 November 2–3; San Antonio, TX. Gen. Tech Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 20–27.

Homer, C.G.; Gallant, A. 2001. Partitioning the conterminous United States into mapping zones for Landsat TM land cover mapping. USGS Draft White Paper.

Kauth, R.J.; Thomas, G.S. 1976. The tasseled cap - a graphic description of the spectral-temporal development of agricultural crops seen by Landsat. In: Proceedings, symposium on machine processing of remotely sensed data. West Lafayette, IN: Laboratory for Applications of Remote Sensing: 41–51.

Lessard, V.; Hansen, M.; Nelson, M. 2003. Comparing Minnesota land cover/use area estimates using NRI and FIA data. In: McRoberts, R.; Reams, G.; Moser, J.; Van Deusen, P., eds. Proceedings, 3d annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 36–43.

Lillesand, T.; Chipman, J.; Nagel, D.; *et al.* 1998. Upper Midwest gap analysis program image processing protocol. Onalaska, WI: U.S. Geological Survey, Environmental Management Technical Center. EMTC 98-G001. 25 p. + Appendixes A–C.

McRoberts, R. 1999. Joint annual forest inventory and monitoring system: the North Central perspective. Journal of Forestry. 97(12): 27–31.

McRoberts, R. 2002. Using maps to increase the precision of forest area estimates. In: Proceedings, 5th International symposium on spatial accuracy assessment in natural resources and environmental sciences; 2002 July 10–12; Melbourne, Australia.

McRoberts, R.; Liknes, G. 2002. Forest/nonforest mapping using inventory data and satellite imagery. In: ForestSat2002: operational tools in forestry using remote sensing techniques; 2002 August 5–9; Edinburgh, Scotland: Forest Research, UK Forestry Commission.

McRoberts, R.; Wendt, D.; Nelson, M.; Hansen, M. 2002a. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of the Environment. 81: 36–44.

McRoberts, R.; Nelson, M.; Wendt, D. 2002b. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sensing of the Environment. 82: 457–468.

Nelson, M.; McRoberts, R.; Wendt, D. (2002). Rapid classification of Thematic Mapper imagery for improving forest inventory and analysis estimates of forest land area. In: Greer, J.D., ed. Proceedings, 9th biennial remote sensing applications conference; 2002 April 8–12 [CD-ROM]. San Diego, CA: U.S. Department of Agriculture, Forest Service. 8 p.

Pouncey, R.; Swanson, K.; Hart, K. 1999. ERDAS field guide™. 5th ed. Atlanta, GA: Erdas®, Inc. 672 p.

U.S. Department of Agriculture. 2002. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.6. Washington, DC: U.S. Department of Agriculture, Forest Service. Internal report. On file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis.

Use of Semivariances for Studies of Landsat TM Image Textural Properties of Loblolly Pine Forests

Jarek Zawadzki[1,4], Chris J. Cieszewski[2], Roger C. Lowe[3], and Michal Zasada[5,6]

Abstract.—We evaluate the applicability of Landsat TM imagery for analyzing textural information of pine forest images by exploring the spatial correlation between pixels measured by semivariances and cross-semivariances calculated from transects of the Landsat TM images. Then, we explore differences in semivariances associated with images of young, middle-aged, and old, and natural versus planted stands. Finally, we compare semivariances for loblolly pine (*Pinus taeda* L.) with those of longleaf pine (*Pinus palustris* Mill.) in Georgia, U.S.A. The results show that, in spite of the low Landsat TM resolution, the semivariances and cross-semivariances may provide useful additional information.

Remotely sensed data are inexpensive supplements to ground measurements and are frequently used in forest inventories of large areas due to the cost efficiency and the ability to provide a large amount of information in a short time (Campbell 1994, Vogelmann *et al.* 1998). Most common methods for image classification of remotely sensed images are applied without considering potentially useful spatial information among various pixels. Semivariograms consider the spatial information and have proved useful in analyzing various spatial data (Curran 1988, Woodcock *et al.* 1988a, 1988b). So far, the semivariograms have been successfully used in forestry applications only with expensive high-resolution data (St.-Onge and Cavayas 1995, Treitz and Howarth 2000).

The objective of our study was to evaluate the applicability of the relatively inexpensive, low-resolution Landsat TM7 TM imagery for analyzing the textural information in images of loblolly pine forests (*Pinus taeda* L.) in Georgia, U.S.A., using geostatistical methods. We analyzed different ages and natural versus planted stands of loblolly pine using semivariograms and cross-semivariograms. To check if semivariograms can discriminate between different species, semivariograms for loblolly pine were compared with those of longleaf pine (*Pinus palustris* Mill.).

We analyzed data from the Thematic Mapper sensor of the Landsat TM7 satellite in combination with ground measurements. We used information from the visible red (RED), the near-infrared (NIR), and the middle-infrared (MIR) bands. The Normalized Difference Vegetation Index (NDVI) as well as the corrected NDVI ($NDVI_c$) and MIR/RED indices were studied.

Area Description, Methods, and Material Studied

Study Site and Data Description

We linked remote sensing images to vegetation data by using data collected in the field. The study area was located in western Georgia, U.S.A. The data collected contained stand information including stand-polygon GIS/GPS coordinates, vegetation type (e.g., species) as well as some quantitative data (e.g., age, basal area, density). We also used data from longleaf pine stands to compare their textural characteristics with another species. We differentiated between planted and natural stands, and divided all stands of both species into three age classes: young (6–11 years), medium (16–26 years), and old (older than 31 years).

Landsat TM data are appropriate for mapping and investigating broad vegetation types classified by the sensor's spectral and spatial characteristics. The important characteristics of the Landsat TM7 satellite are:

[1,5]Postdoctoral Fellow, [2]Assistant Professor, [3]GIS Analyst, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602, USA, biomat@uga.edu.
[4]Assistant Professor, Environmental Engineering Department, Warsaw University of Technology, Nowowiejska 20, 00–61 Warsaw, Poland, jarek97@yahoo.com.
[6]Assistant Professor, Department of Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, Rakowiecka 26/30, 02–528 Warsaw, Poland, zasada@delta.sggw.waw.pl.

1. scene coverage–115 miles by 115 miles
2. spectral resolution–three bands in the visible portion of the spectrum, three bands in the reflective-infrared portion of the spectrum, one band in the thermal portion of the spectrum, and a panchromatic (black and white) band
3. spatial resolution–30 meters for the visible band
4. temporal resolution–16 days

We used digital numbers (DN) from the RED band (red, 0.63-0.69$_{\mu}$m), the NIR band (reflective-infrared, 0.76-0.90$_{\mu}$m), and the MIR band (mid-infrared 1.55-1.75$_{\mu}$m). The RED band is sensitive enough for discriminating between many plant species. The NIR band is especially sensitive to the amount of vegetation biomass present in a scene. The MIR band is sensitive to the amount of water in plants (ERDAS Field Guide 1990). Finally, we also studied the geostatistical characteristics of the Normalized Difference Vegetation Index (NDVI) by Rouse (1973), and the corrected $NDVI_c$ as well as the MIR to RED ratio vegetation index (MIR/RED) by Jordan (1969). The NDVI was calculated according to the following formula:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

where RED and MIR denotes the red and the near-infrared reflectance. The $NDVI_c$ vegetation index is a NDVI modified index, especially designed for distinguishing coniferous forests (Nemani *et al.* 1993). $NDVI_c$ is given:

$$NDVI_c = \frac{MIR - RED}{MIR + RED}\left(1 - \frac{NIR - NIR_{min}}{NIR_{max} - NIR_{min}}\right)$$

where the first factor in the equation is the NDVI and the second factor is a correction of the NDVI. The NIR_{min} is the reflectance value of pixels corresponding to field plots with the lowest tree canopy, and NIR_{max} is the reflectance value of pixels with the highest canopy cover.

All remotely sensed images were analyzed using ERDAS Imagine 8.5 Software.

Methods

Geostatistics comprises many methods for evaluating the autocorrelation that commonly exists in spatial data. The main tool of geostatistics is the semivariogram (semivariance), which is a measure of spatial continuity. The experimental semivariogram is derived by calculating half the average squared difference in data values for every pair of data locations along a specified direction:

$$\gamma = \frac{1}{2N}\sum_{i=1}^{N}\left[Z(x_i) - Z(x_i + h)\right]^2$$

where x_i is a data location, h is a lag vector, $Z(x_i)$ is the data value at location x_i, N is the number of data pairs spaced a distance and direction h units apart. These values are then plotted against the distances between data pairs. Such a plot is commonly referred to as a variogram and has a classic form shown in figure 1.

Semivariograms are roughly defined by three characteristics:

1. sill–the plateau that the semivariogram reaches. The sill is the amount of variation explained by the spatial structure.
2. range of the influence (correlation)–the distance at which the semivariogram reaches the sill.
3. nugget effect–the vertical discontinuity at the origin. The nugget effect is a combination of sampling error and short-scale variations that occur at a scale smaller than the closest sample spacing. The sum of the nugget effect and the sill is equal to the variance of the sample.

Figure 1.—*The "classic" form of semivariance.*

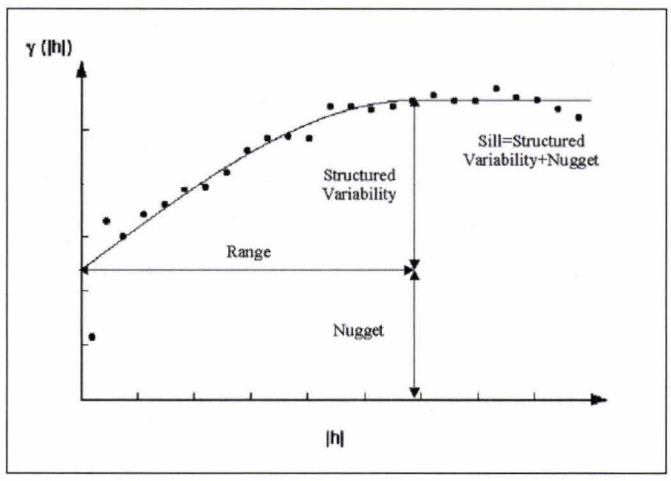
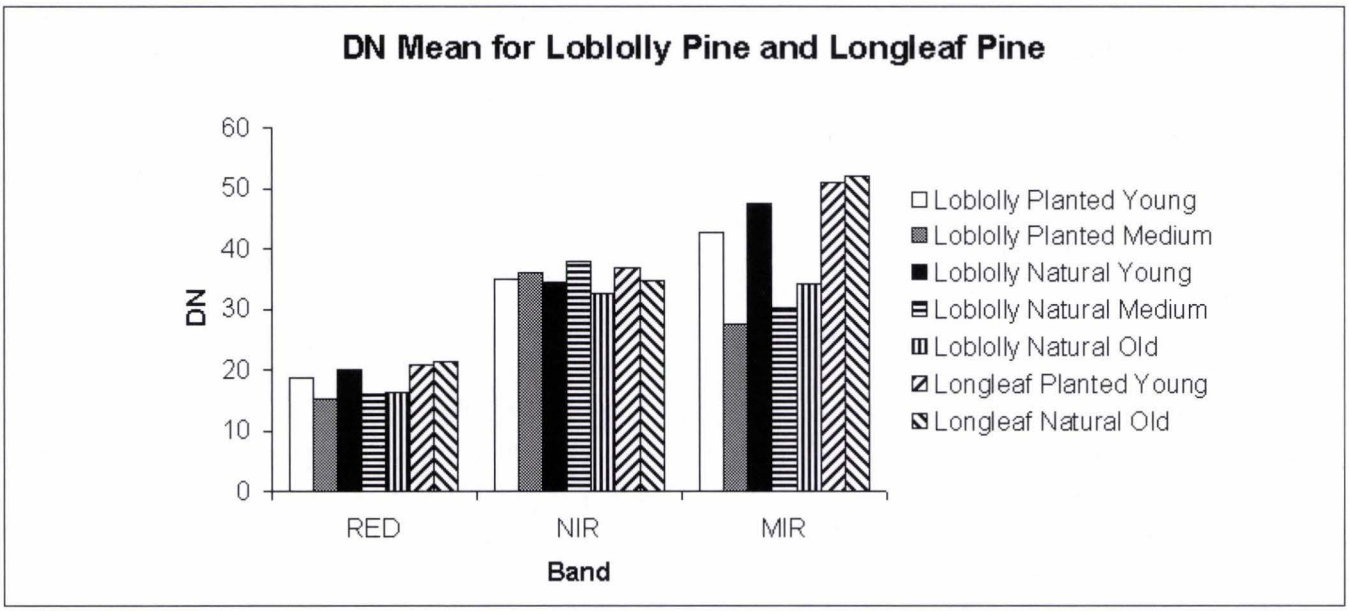


Figure 2.—*Mean of DN for loblolly pine and longleaf pine calculated from Landsat TM image, channels RED, NIR and MIR, stand origin 1988.*



The obtained experimental semivariogram is used to fit an appropriate theoretical model, as e.g., spherical, exponential, etc., and can be used in other geostatistical analyses, e.g. kriging.

Remotely sensed images can be also used in semivariogram calculations. The semivariogram is calculated from the transects running across a remotely sensed image using digital numbers as data values $Z(x_i)$.

Another important measure of spatial correlation is the cross-semivariogram:

$$\gamma_{WZ} = \frac{1}{2N}\sum_{i=1}^{N}\left[W(x_i) - W(x_i + h)\right]\left[Z(x_i) - Z(x_i + h)\right]$$

where x_i is a data location, h is a lag vector, $Z(x_i)$ and $W(x_i)$ are the DN values at location x for different bands, N is the number of data pairs spaced a distance and direction h units apart. The cross-semivariogram quantifies the joint spatial variability (cross correlation) between two radiometric bands.

Semivariograms can be a useful tool in classification, but there are some important difficulties in applying semivariograms to forest classification. First of all, often in forested areas semivariograms are much more complicated than the "classic" ones. For example, some periodic and aspatial variations of the classic semivariogram were often observed for forested areas. The first type of semivariogram appears when a repetitive pattern is studied, and the second one appears when random patterns are investigated. There were also "unbounded" forms of semivariograms observed in the study. The unbounded semivariogram may represent a situation in which a trend or many spatially correlated phenomena exist. These nonclassic semivariograms are much more difficult to model and interpret.

Results and Analyses

The basic descriptive statistics of the analyzed forest types (fig. 2) reveal some distinctions between the different stands but do not provide any textural information. To explore the textural continuity of studied stands, we calculated and analyzed the semivariograms for RED, MIR, and NIR bands, as well as the cross-semivariograms between these bands. The semivariograms for the above mentioned vegetation indices were also calculated.

To understand better the factors that influence the semivariograms, we calculated them in large and potentially homogeneous areas, changing for comparison only one essential stand feature, e.g., age (young, medium, old) or type of stand (planted, natural). Figure 3 shows typical, standardized (divided by theirs variances)

Figure 3.—*Isotropic semivariances for a planted loblolly pine stand calculated from Landsat TM images; stand origin 1988.*

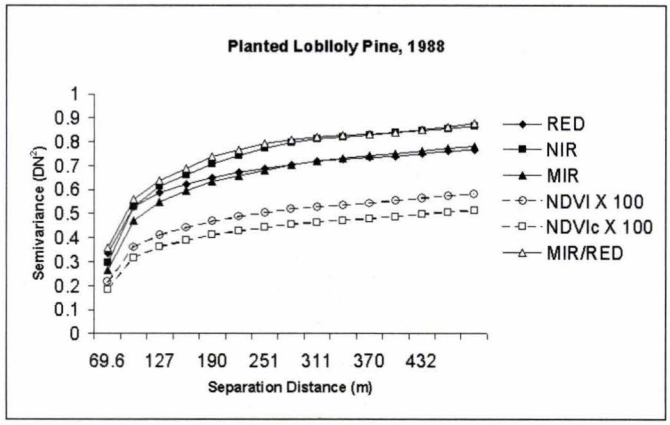
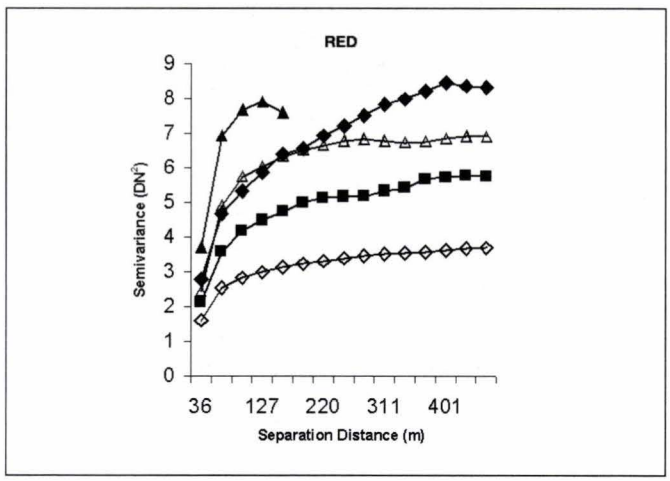


Figure 4.—*Isotropic semivariances for different types of loblolly pine stands, calculated from Landsat TM image using RED band.*



semivariograms, calculated for 12-year-old planted loblolly pine stands. These semivariograms have a typical "unbounded" shape showing many spatially correlated phenomena.

For the small separation distances (a few lags) the semivariogram curve rises relatively fast. Then, at the greater distances it exhibits a gentle sloping and becomes almost linear. The initial increase of the semivariogram curve results from the fast decrease of spatial continuity at the distances of a few lags (1 lag = 30 meters). This means that the spatial correlations between pixels decreases rapidly for short distances.

At longer distances, the semivariograms do not reach saturation but increase almost linearly. This means that many sizes and shapes of the forest stands are present in the scene. As was already mentioned, these semivariograms are difficult to use for classification purposes. For example, it is clear that the range and the sill cannot be distinctive parameters for different vegetation communities (fig. 3).

To check whether the semivariograms can be treated as "spatial signatures" of different type of coniferous forests, we calculated them for different types of loblolly pine stands. We calculated semivariograms using the DNs from RED, NIR, and MIR bands as well as NDVI and MIR/RED indices. The largest differences between semivariograms calculated for the investigated loblolly pine stands were obtained from the RED and MIR bands. The results of the calculations for the RED band are shown in figure 4.

Distinctly smaller differences were observed between semivariograms calculated for DN from the NIR band as well as between semivariograms calculated from the vegetation indices NDVI and MIR/RED. This somewhat surprising behavior of semivariograms from vegetation indices can be explained by the smoothing effect; these indices are the ratios of DNs coming from different bands.

Natural stands have higher semivariogram values than even-aged planted stands (fig. 4). This is because natural stands' have a higher textural variability than planted stands.

We also compared semivariograms for different species of pine by calculating semivariograms for planted and natural stands of longleaf pine. The exemplary semivariograms of loblolly pine and longleaf pine calculated from the DN for the MIR band are shown in figure 5.

Large differences exist between semivariograms calculated from loblolly pine and longleaf pine stands. The semivariogram values for longleaf pine are much higher than those of loblolly pine, calculated for the stands of similar type and age. The values of semivariograms at the distance of a few lags can be also used as a discriminative parameter.

The cross-semivariograms quantify the joint spatial variability between two bands. Therefore, they can be also used for texture-based classification adding new spatial information. So, at the end of our analysis we calculated also cross-semivariograms between bands RED, MIR, and NIR for planted and natural, medium-aged stands of loblolly pine. The largest cross-correlations were between the RED and MIR bands both for the planted and for natural stands. The cross-correlations between bands RED and NIR as well as between MIR and NIR were substantially smaller. The values of the cross-semivariogram for natural stands were much higher than for planted stands. The cross-cor-

relations between the RED and MIR bands calculated for studied loblolly pine stands are shown in figure 6. Clearly, all age classes are well separated. The largest cross-semivariogram values were obtained for young stands and the smallest for old stands, both for planted and natural stands.

Conclusions

In spite of the low-resolution of the remote imagery, distinct differences were found in semivariograms of images for the studied forests. This means that such semivariograms can be treated as "spatial signatures" for the studied forest stands. The classical semivariogram's parameters, such as range and sill, are not appropriate as differentiated parameters because of the low-resolution of the remote imagery and the nonclassic, unbounded type of observed semivariograms. However, there are important differences for semivariogram and cross-semivariogram values at the distances of several lags. Our study suggests that the semivariogram values for such separation distances (e.g., from the 4th to the 7th lags) are appropriate for these purposes. The observed differences between semivariograms at distances of several lags arise from different spatial correlations existing in the studied forest stands at distances of a few tens to a few hundred meters. The low-resolution of Landsat TM7 remote imagery does not allow distinguishing separate trees. The observed spatial correlations can be attributed to the similarity in arrangements of bigger objects as groups of trees (or stands), areas with similar underbrush, etc. The largest differences in semivariograms were obtained for RED and MIR bands.

Figure 5.—*Semivariograms calculated from remote images of loblolly pine and longleaf pine stands using MIR band.*

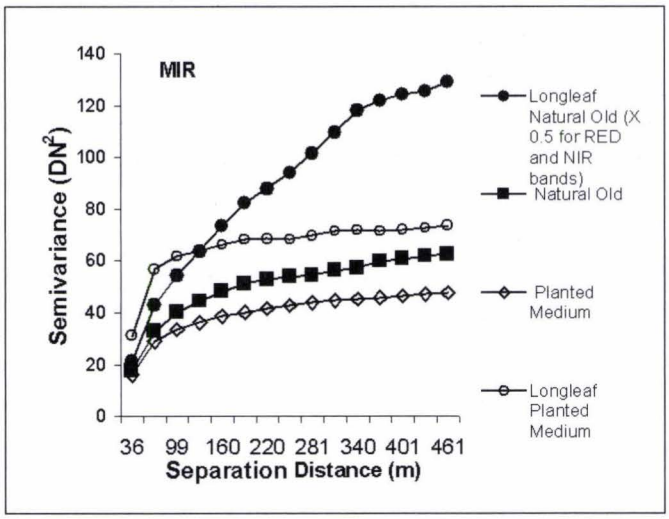
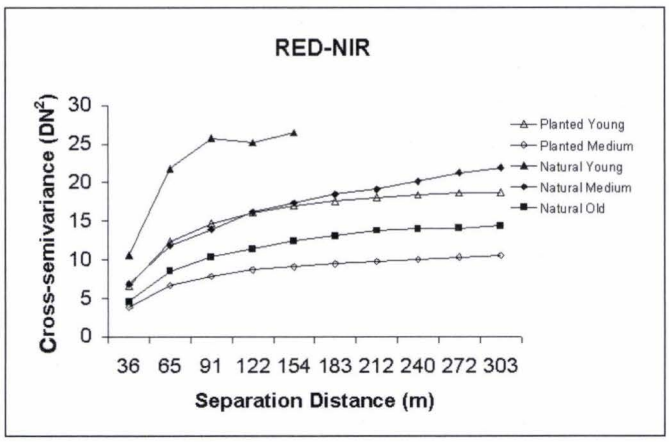


Figure 6.—*The cross-correlations between the RED and NIR bands calculated for studied loblolly pine stands.*



Acknowledgments

This study was sponsored by the School of Forest Resources, University of Georgia, and the Traditional Industry Program 3 (TIP3) funding. We are much indebted for this support. Jaroslaw Zawadzki thanks the Warnell School of Forest Resources, University of Georgia, for a creative and hospitable environment; Ingvar Elle for help with English; and Tripp Lowe for technical assistance.

Literature Cited

Campbell, J.B. 1994. Introduction to remote sensing. New York, NY: Guilford Press.

Curran, P.J. 1988. The semivariances in remote sensing: an introduction. Remote Sensing of Environment, 24: 493–507.

ERDAS, Inc. 1990. ERDAS field guide, 5th ed. 1990. Atlanta, GA: ERDAS, Inc.

Jordan, C.F. 1969. Derivation of leaf area index from quality of light on the forest floor. Ecology. 50: 663–666.

Nemani, R.; Pierce, L.; Running, S. 1993. Forest ecosystem process at the watershed scale: Sensitivity to remotely-sensed leaf area index estimates. International Journal of Remote Sensing. 14: 2519–2539.

Rouse, J.W.; Hass. R.H.; Schell, J.A.; Deering, D.W. 1973. Monitoring vegetation systems in the great plains with ERTS. In: Proceedings, 3d ERTS symposium; [Dates held unknown]; [where held unknown]. NASA SP–351. Washington, DC; NASA: 309–317.

St.-Onge, B.A.; Cavayas, F. 1995. Estimating forest stand structure from high-resolution imagery using the directional semivariogram. International Journal of Remote Sensing. 16: 1999–2001.

Treitz, P.; Howarth, P. 2000. High spatial remote sensing data for forest ecosystem classification: an examination of spatial scale. Remote Sensing of Environment. 72: 268–289.

Vogelmann, J.E.; Sohl, T.L.; Campbell, P.V.; Shaw, D.M. 1998. Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources. Environmental Monitoring and Assessment. 51: 415–428.

Woodcock, C.E.; Strahler, A.H.; Jupp, D.B. 1988a. The use of semivariograms in remote sensing: I: scene models and simulated images. Remote Sensing of Environment. 25: 323–348.

Woodcock, C.E.; Strahler, A.H.; Jupp, D.B. 1988b. The use of semivariograms in remote sensing: II: real digital images. Remote Sensing of Environment. 25: 349–379.

Alternative Natural Resource Monitoring Strategies in the Mexican States of Jalisco and Colima

Cele Aguirre-Bravo and Hans Schreuder[1]

Abstract.—This paper presents a strategy for inventorying and monitoring the natural resources in the Mexican states of Jalisco and Colima. The strategy emphasizes a strong linkage between remote sensing with field sampling design to produce statistical summaries and spatial estimates at multiple scales and resolution levels. Outputs derived from this strategy are expected to have significant local use where policy and management decisions are most effectively made.

Through the Consortium for Advancing the Monitoring of Ecosystem Sustainability in the Americas (CAMESA), various federal agencies of Canada, the United States, and Mexico are working in partnership with the Mexican states of Jalisco and Colima to advance the science and technology of monitoring natural resource sustainability at multiple scales and resolution levels. Central to this work is the need to design and implement inventory and monitoring programs that are cost effective, technically defensible, scientifically credible, and of high social utility for multiple natural resource applications. Within the USDA Forest Service, the Rocky Mountain Research Station is currently leading this effort. In light of the successful results so far, these cooperative activities have been organized into a **Pilot Study and Learning Center** so that a variety of stakeholders (i.e., federal and State government, industry, academia, and nongovernmental organizations) can learn and benefit from the results of this experience. Should this undertaking succeed, it could then be recommended as a model for implementing similar initiatives elsewhere in Mexico and other countries in the Americas.

Vision and Mission

The vision in this cooperative undertaking is to advance the human values of environmental sustainability and the socioeconomic system by providing the resources that will enable individuals, agencies, organizations, governments, and other entities in Mexico and the Americas to effectively manage their natural resources in a way that sustains, rather than degrades, the ecosystem and enriches, rather than impoverishes, the social and cultural environment.

Goal and Objectives

The pilot study's overall goal is to develop a well-structured set of strategies that incorporates state-of-the-art science, technology, and analytical capabilities together with an Internet-based strategy for communicating program outputs. A seamless process will be created to provide all stakeholders the tools and knowledge needed to make intelligent decisions regarding the profitable and sustainable management and utilization of natural resources. Investments in this program will be directed to achieving the following specific objectives:

- Train a cadre of workers to inventory, monitor, and assess the sustainable management of natural resources.
- Create an appropriate physical infrastructure through which users may acquire, store, manage, analyze, report, and disseminate the information to inventory and monitor natural resources for their profitable and sustainable management and utilization.
- Create strategies for developing accurate and useful information about natural resources to help in decisionmaking and planning for sustainability.

If successful, this cooperative initiative will facilitate a multi-institutional capability for managing natural resources

[1] Research Coordinator for the Americas and Mathematical Statistician, respectively, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2150 Centre Ave., Building A, Fort Collins, CO 80526–1891; e-mail: caguirrebravo@fs.fed.us, hschreuder@fs.fed.us.

for their sustained use and productivity, while at the same time ensuring their vitality, diversity, and ability to provide important ecological services for the enjoyment of present and future generations.

Scope and Context

While focused on the goals described above, and using a hierarchical system of "Watershed Units" for ecological and economic accounting, the program will address a variety of critical questions regarding the information needed for assessing and managing natural resources within watersheds, at multiple scales and resolution levels. For example, what are the extent and condition of the watershed resources and processes (i.e., vegetation, soils, water, animals, landscapes, runoff, erosion, human activity, etc.)? What components of the watershed are changing and why? Why are some resources changing faster than others and where are these changes taking place (i.e., within and across watersheds)? What are the quantity, quality, and extent of services provided by watersheds, and how do human systems benefit from them? Within and across watersheds, where is mitigation/restoration of resources and processes most practical and beneficial? How are human systems sustaining the ecological integrity and societal value of watersheds? Will the current extent and condition of resources/services of watersheds meet future ecological and economic needs? How can stakeholders (i.e., landowners, federal and State government, industry, academia, and nongovernmental organizations) work together to solve the problems and issues within and across watersheds so that we can ensure the health of these systems and the well-being of present and future generations? Specific strategies will be developed and implemented to address these questions.

Study Area Significance

The Pilot Study Area comprises the Mexican southwestern states of Jalisco and Colima. The two states together cover an area of approximately 10 million hectares (25 million acres). Although Jalisco occupies 90 percent of the area, the State of Colima plays a distinctive role in the economy of the whole region and helps to diversify the Pilot Study Area. Four major ecological regions provide the natural resources and environmental conditions that make this region one of the most prosperous in Mexico. These ecoregions are the transversal neo-volcanic system, the southern Sierra Madre, the Southern and Western Pacific Coastal Plain and Hills and Canyons, and the Mexican High Plateau. Nested within these ecological regions are several important Hydrological Regions (HR) that drain to the Pacific Ocean: (HR12 Lerma-Santiago, HR13 Huicicila, HR14 Ameca, HR15 Costa de Jalisco, HR16 Armeria-Coahuayana, HR18 Balsas, and HR37 El Salado). One of the watersheds, the Lerma-Santiago Hydrological Region, is connected to Chapala Lake, the primary source of water for the City of Guadalajara.

Variables and Indicators

To maximize data versatility, the variables and indicators proposed in this pilot study either are directly parallel or are similar to those used in inventory and monitoring programs used by land management and environmental protection agencies of the United States and Canada. By adopting variables and indicators that meet Quality Assurance/Quality Control (QA/QC) requirements, the data and information collected through this program should be fully compatible or comparable with North American databases and other international programs. Local needs for more specific information for natural resource management will influence significantly the collection of additional variables and indicators. Long-term comparability of variables and indicators between and among various jurisdictions can be achieved by measuring a minimum subset of them **(Core Variables and Indicators)** at each site to address issues and problems of common concern. Some of the core variables will be standard information necessary to locate each site and define its basic physical and resource use characteristics. These variables will be measured once or remeasured occasionally. Other "Core Variables and Indicators" will also be measured to document the factors that have historically affected the status of the ecosystem. Metadata records will be kept to document these processes.

Strategy for Sampling Design

In designing an integrated multi-resource inventory and monitoring system to evaluate the condition and change of variables and indicators for sustainable natural resource management (forest, rangeland, agriculture, wildlife, water, soils, biodiversity, etc.), one needs some baseline data for comparison. Because one is generally dealing with complex systems, it may not be wise to select one or two variables for ecological monitoring. Also, analyzing these variables independently may lead to incorrect conclusions because of their interdependencies. One approach is to model the spatial relationship between key indicator variables. In natural resource management, for example, this information can be used to identify forest habitats that are either conducive or a deterrent to the presence of ecologically important plant and/or animal species. Techniques commonly used in describing spatial relationships between two or more variables include regression analysis and a variety of spatial statistical procedures that take into consideration the spatial dependency. The proposed natural resource monitoring system will rely on information collected at different spatial scales of resolution and sampling intensities to provide detailed information at the local level for natural resource planning and management (Schreuder *et al.* 2002).

Spatially Continuous Monitoring

Landsat Thematic Mapper (TM) data will be used to provide a complete and uniform census of individual Environmental Accounting Units (EAU) across jurisdictions (i.e., private lands, federal and state lands, ejidos, communities, counties, regions, etc.). This approach will provide measurements collected as a series of contiguous and simultaneous measures across land tenure units. It will also allow monitoring EAU's for changes in spectral and spatial characteristics that can be applied over a range of spatial and temporal scales appropriate for addressing specific natural resource issues.

Design-based Monitoring

The development of the sampling and plot designs is complicated by the variety of indicators to be assessed, the need to assess the natural resources at a range of scales, the need to monitor the indicators over time, and the need to do so efficiently. To meet national and State level objectives for natural resource assessments, we will develop a grid-based, traditional sampling design. For the remaining objectives involving estimation at local scales, the design will be enhanced to provide information needed to develop spatial statistical models to estimate key attributes at all locations within the sampled population (Reich and Aguirre-Bravo 2002).

Site-specific Monitoring

Because of the biological importance of certain areas in terms of threatened and endangered plant and animal species, there is a great need to initiate species- and/or site-specific research and monitoring. Detailed information should be collected at this stage through specific projects if funding is available. In addition, data should be collected that are compatible with key indicator variables collected at other monitoring levels. All data should be georeferenced to allow the integration with information collected at the different levels using spatially explicit models.

Quality Assurance/Quality Control

Quality assurance (QA) and quality control (QC) are essential to any monitoring and inventorying system. Quality Assurance builds confidence in the results of the inventory and monitoring program. Quality Control documents the quality of various program components to ensure that the components meet some minimal level of desired quality. QA/QC protocols apply to all components of the inventory program: program planning, data collection, information management and compilation, analysis and reporting, and continuous program improvement. Compatibility of QA/QC protocols across monitoring programs in the NAFTA countries is essential for this pilot study (Aguirre-Bravo and Alonso 2002).

Statistical Spatial Modeling

Detailed spatial models describing natural resources are typically limited by the spatial resolution of the data on which the models are based. For large geographical surveys, obtaining sufficient coverage further complicates the modeling process, as time and resource limitations preclude detailed sampling over large areas. Use of remotely sensed information, such as

multi-spectral satellite imagery, allows one to easily derive large amounts of resource information over large areas; however, these sensors gather information at a fixed spatial resolution (e.g., 30 x 30 m for Landsat satellite imagery), and resources (such as forest structure, soil properties, etc.) may still occur at scales smaller than the resolution of the sensors with which the data were collected. Both Metzger (1997) and Joy and Reich (2002), however, were able to improve the spatial resolution of satellite-based classifications of forest structure and composition, respectively, using fine-scale field data for the classification procedure.

Modeling key indicator variables will be similar to what was done in Metzger (1997) and Joy and Reich (2002). Ordinary least squares (OLS) procedures will be used to generate trend surface models (TS) that describe the large-scale spatial variability in each of the vegetative elements measured in the field. Proportional data (basal areas by species, canopy closure, etc.) will be transformed using logistic transformation to stabilize the variance of the large and small proportions. OLS will be used twice in the model-building process —once as a preliminary means to reduce the number of independent variables used to predict the vegetative characteristics of interest, and secondly to generate the final TS model. In the preliminary analysis, independent variables with a P-value > 0.15 will be dropped. Independent variables used in the model may include slope, aspect, elevation, landform, information from Landsat bands 1-5 and 7, and land use class. Dummy variables can be added to the TS models to account for interactions between the various land use classes and the other independent variables.

Combinatorial regression will be used to determine which of the remaining independent variables best predict the dependent variable of interest. This "screening" procedure, which examines all possible combinations of independent variables in all possible orders to yield the best fit, determined by the lowest Akaike's Information Criteria (Akaike 1973), requires enormous amounts of computer memory. So it is important to eliminate unnecessary variables in the model (hence, the preliminary OLS) before running this screening procedure. The best fitting models will be used to generate a grid for each structural component of interest using ARC/INFO® (ESRI 1995). Kriging, cokriging, or regression trees will be used to describe small-scale spatial variability (i.e., error associated with the residuals from each TS model) in the landscape (Reich and Aguirre-Bravo 2002).

Project Coordination

Initially, the proposed organizational structure for coordinating this undertaking consists of a Project Technical Coordinator, a Science and Technical Committee (STC), and various Task Force Units for technical training, field implementation, data analysis, information management, and reporting. Members of the STC are senior executives/scientists from CAMESA institutional partners, as well as experts from other participating institutions and organizations. In this organization, the STC provides a mechanism to foster and coordinate technical and scientific cooperation and collaboration on matters concerning the design, planning, and execution of activities related to this project. Periodic meetings with the STC serve to analyze strategies and recommend ways to successfully implement this project. Aguirre-Bravo and Alonso (2002) provide detailed information about the coordination, organization, and implementation of this pilot study project.

Expected Products and Benefits

The pilot study creates a window of opportunity for a coordinated multinational effort to design and implement integrated approaches for inventorying and monitoring natural resources in the Mexican states of Jalisco and Colima. Salient to this undertaking is the opportunity to further improve information compatibility and procedures for use in integrating and evaluating information on the status, extent, trends, and projected changes in natural resources across jurisdictions, and at multiple scales and resolution levels. As a multinational partnership effort, it promotes the sharing of scientific and technical information and approaches to gain common understanding on a variety of issues and problems of current and future concern within and across jurisdictional boundaries and geographical scales. In addition, it addresses new approaches and methodologies for advancing the design and implementation of inventory and monitoring programs for the assessment and

sustainable management of natural resources, at multiple scales and resolution levels. For the Americas, and particularly for Mexico, the pilot study serves as a learning center upon which scientists and resource managers learn and benefit from the results of working in partnership.

Acknowledgment

The authors thank the Rocky Mountain Research Station, FIA program, International Forestry of the USDA Forest Service, as well as the State of Jalisco Foundation for Forestry Development (FIPRODEFO) for their cooperation in this study.

Literature Cited

Aguirre-Bravo, C.; Alonso, T.L.A. 2002. Conceptual framework for inventorying and monitoring ecosystem resources at multiple scales and resolution levels. Tech. Bull. 35. PRODEFO, State Government of Jalisco, Mexico: 33–61.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petran, B.N.; Csaki, F., eds. International symposium on information theory, 2d ed. Budapest, Hungary: Akademiai Kiadi: 267–281.

ESRI. 1995. ARC/INFO® software and on-line help manual. Redlands, CA: Environmental Research Institute, Inc.

Joy, S.M.; Reich, R.M. 2002 [In review]. Modeling forest structure on the Kaibab National Forest in Arizona. Forest Science.

Metzger, K. 1997. Modeling forest stand structure to a ten meter resolution using Landsat TM data. Fort Collins, CO: Colorado State University. 123 p. M.S. thesis.

Reich, R.M.; Aguirre-Bravo, C. 2002. Spatial statistical strategy for inventorying and monitoring ecosystem resources. Tech. Bull. 35. Mexico: PRODEFO, State Government of Jalisco. 121–130.

Schreuder, H.; Williams, M.S.; Aguirre-Bravo, C.; Patterson, P.L. 2002. Statistical strategy for inventorying and monitoring the ecosystem resources of the states of Jalisco and Colima at multiple scales and resolution levels. Tech. Bull. 35. Mexico: PRODEFO, State Government of Jalisco. 87–108.

Coordination, Cooperation, and Collaboration between FIA and NRI

Raymond L. Czaplewski[1], James Rack[2], Veronica C. Lessard[3], David F. Heinzen[4], Susan Ploetz[5], Thomas L. Schmidt[6], and Earl C. Leatherberry[7]

Abstract.—The USDA Forest Service conducts a detailed survey of the Nation's forests through the Forest Inventory and Analysis (FIA) program. The USDA Natural Resources Service conducts an entirely separate survey, the National Resources Inventory (NRI), to monitor status and trends in the Nation's soil and other natural resources. Blue Ribbon Panels for both FIA and NRI have recommended better cooperation and collaboration. In response, a joint venture among the State of Minnesota, the U.S. Geological Survey, NRI, and FIA searched for potential synergies by fusing FIA and NRI plot data with Landsat imagery and a statewide geographic information system. FIA and NRI plot data did prove useful as training data for classifying land cover, and as supplemental labeling data for detecting changes with multi-date Landsat imagery.

The U.S. Department of Agriculture (USDA) conducts three statistical surveys of the Nation's natural resources:

1. The USDA National Agricultural Statistics Service (NASS) estimates annual production and supplies of food and fiber, prices paid and received by farmers, farm labor and wages, and farm aspects of the agricultural industry (e.g., pesticide use). The annual NASS budget is approximately $100 million.
2. The National Resources Inventory (NRI) is conducted by the USDA Natural Resources Conservation Service (NRCS) on all non-Federal lands. NRI estimates the extent of different kinds of land cover and land use in the USA; indicators of soil condition and erosion; and the extent and changes in land management; wetlands; and other natural resources. For example, NRI estimates area of cropland, pastureland, rangeland, land enrolled in the Conservation Reserve Program, other rural land, builtup and urban land, water bodies, and forestland (including nonstocked and 22 broad categories of forest type).
3. The Forest Inventory and Analysis (FIA) program is conducted by the USDA Forest Service. FIA estimates tree, site, and stand conditions of the Nation's forests. For example, FIA estimates the area of forestlands by many detailed categories of stand conditions. The FIA budget was $49 million in 2001, with an additional $8 million in State funds (USDA 2002).

Each of these USDA surveys is well designed to implement a different congressional mandate that relates to the inventory of natural resources. Each mandate serves a distinct group of customers, each with its own unique blend of natural resource issues. Each survey uses its own sampling designs,

[1] Project Leader, Rocky Mountain Research Station, U.S. Department of Agriculture, Forest Service, Fort Collins, CO 80526. Phone: 970–295–5973; fax: 970–295–5959; e-mail: rczaplewski@fs.fed.us.

[2] Senior Research Analyst, Resource Assessment Unit, Minnesota Department of Natural Resources, Division of Forestry, Grand Rapids, MN. Phone: 218–327–4449; fax: 218–327–4517; e-mail: jim.rack@dnr.state.mn.us.

[3] Statistician, Natural Resources Inventory and Analysis Institute, U.S. Department of Agriculture, Natural Resources Conservation Service, co-located with the North Central Research Station, U.S. Department of Agriculture, Forest Service, St. Paul, MN 55108. Phone: 651–649–5130; fax: 651–649–5140; e-mail: vlessard@fs.fed.us.

[4] Supervisor, Resource Assessment Unit, Minnesota Department of Natural Resources, Division of Forestry, Grand Rapids, MN. Phone: 218–327–4449; fax: 218–327–4517; e-mail: david.heinzen@dnr.state.mn.us.

[5] Soil Conservationist, St. Paul State Office; U.S. Department of Agriculture, Natural Resources Conservation Service, St. Paul, MN 55108. Phone: 651–602–7888; fax: 651–602–7914; e-mail: Susan.Ploetz@usda.gov.

[6] Assistant Director for Research, North Central Research Station, U.S. Department of Agriculture, Forest Service, St. Paul, MN 55108. Phone: 651–649–5131; fax: 651–649–5285; e-mail: tschmidt@fs.fed.us.

[7] Research Social Scientist, North Central Research Station, U.S. Department of Agriculture, Forest Service, St. Paul, MN 55108. Phone: 651–649–5138; fax: 651–649–5140; e-mail: eleatherberry@fs.fed.us.

protocols, and definitions that are designed to best serve its own mission.

Unfortunately, differences among USDA surveys create discrepancies among a few important variables that overlap surveys, such as area of forestland. For example, NRI estimates for acres of forest can differ from FIA estimates by over 30 percent (fig. 1). Why? FIA and NRI define urban and builtup lands differently. They can use different sources and dates of administrative data to develop area expansion factors. While both FIA and NRI define forest to be at least 10 percent stocked, this definition is applied with different protocols. NRI classifies some land with forest cover as Conservation Reserve Program (CRP), while FIA classifies the same lands as forest. In Minnesota, NRI often classifies as forest the tall shrubland within the transition zone between forest and inland marshes and swamps, while FIA classifies the same areas as nonforest. FIA classifies vast areas of oak, pinyon, and juniper woodland as forest in the interior west, while NRI often classifies the same areas as shrubland or rangeland.

In 1998, a team of senior scientists from the FIA, NRI, NASS, U.S. Geological Survey (USGS), Bureau of Land Management, and Environmental Protection Agency demonstrated the feasibility of combining FIA and NRI surveys while preserving critical historic information (House *et al.* 1998, USDA 1998b). They formulated a framework for estimating the extent of forest and rangeland that explains the discrepancies between FIA and NRI estimates. This framework envisioned a joint USDA inventory and monitoring effort for terrestrial natural resources that links the FIA and NRI surveys through a co-located subset of sample plots and a shared database.

Figure 1.—*Differences between NRI and FIA in estimated number of acres of forest. Discrepancies are primarily caused by differences in measurement protocols and definitions for land cover v. land use.*

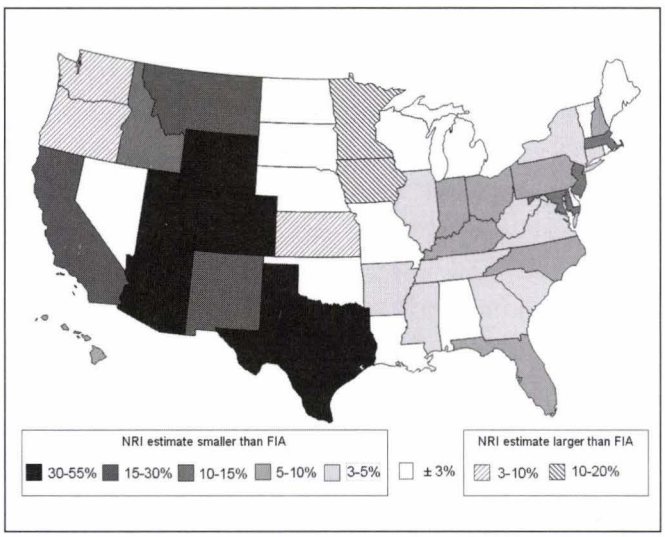
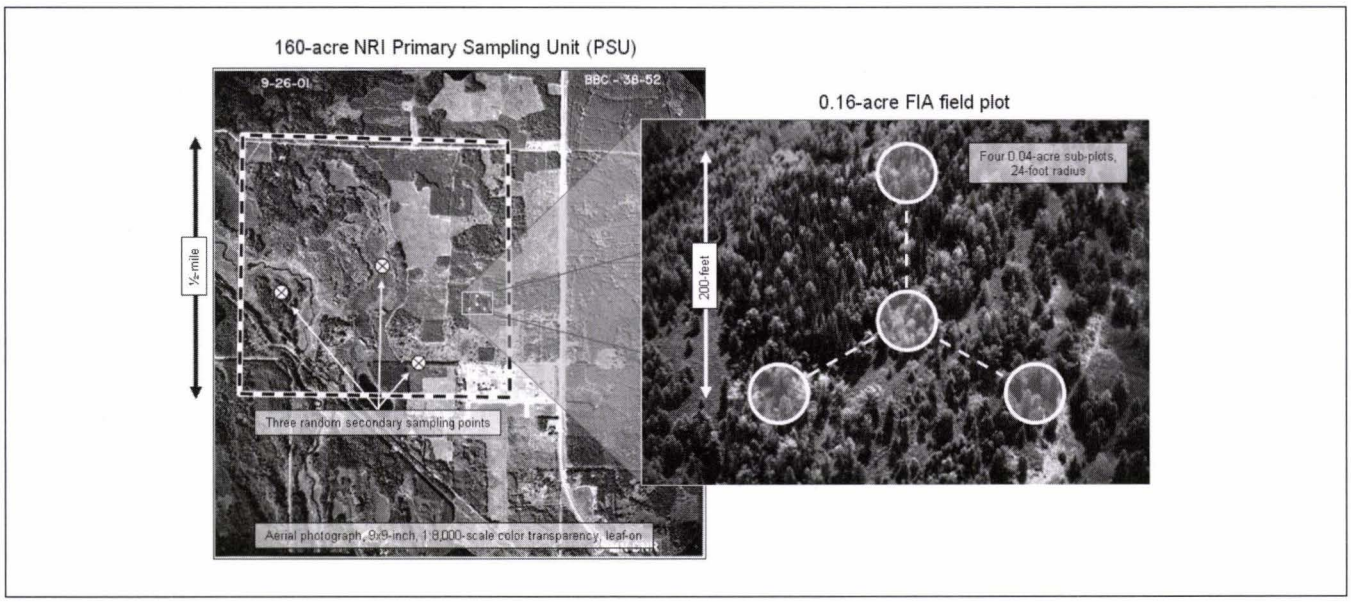


Figure 2.—*Comparison of a 160-acre NRI Primary Sampling Unit with a FIA 1-acre field plot. The NRI 1:8,000-scale aerial photographs encompass approximately 5 percent of the landscape; therefore, only about 5 percent of FIA field plots are imaged within NRI sample photographs. These are demonstration plots, and they are not part of the FIA or NRI sampling frames.*



Beginning the following year, a second team of scientists investigated a fusion of the independent databases produced by FIA and NRI within a geographic information system (GIS), without a shared subset of co-located sample plots. This paper briefly summarizes the results of these latter experiments and suggests future experiments to improve collaboration between FIA and NRI.

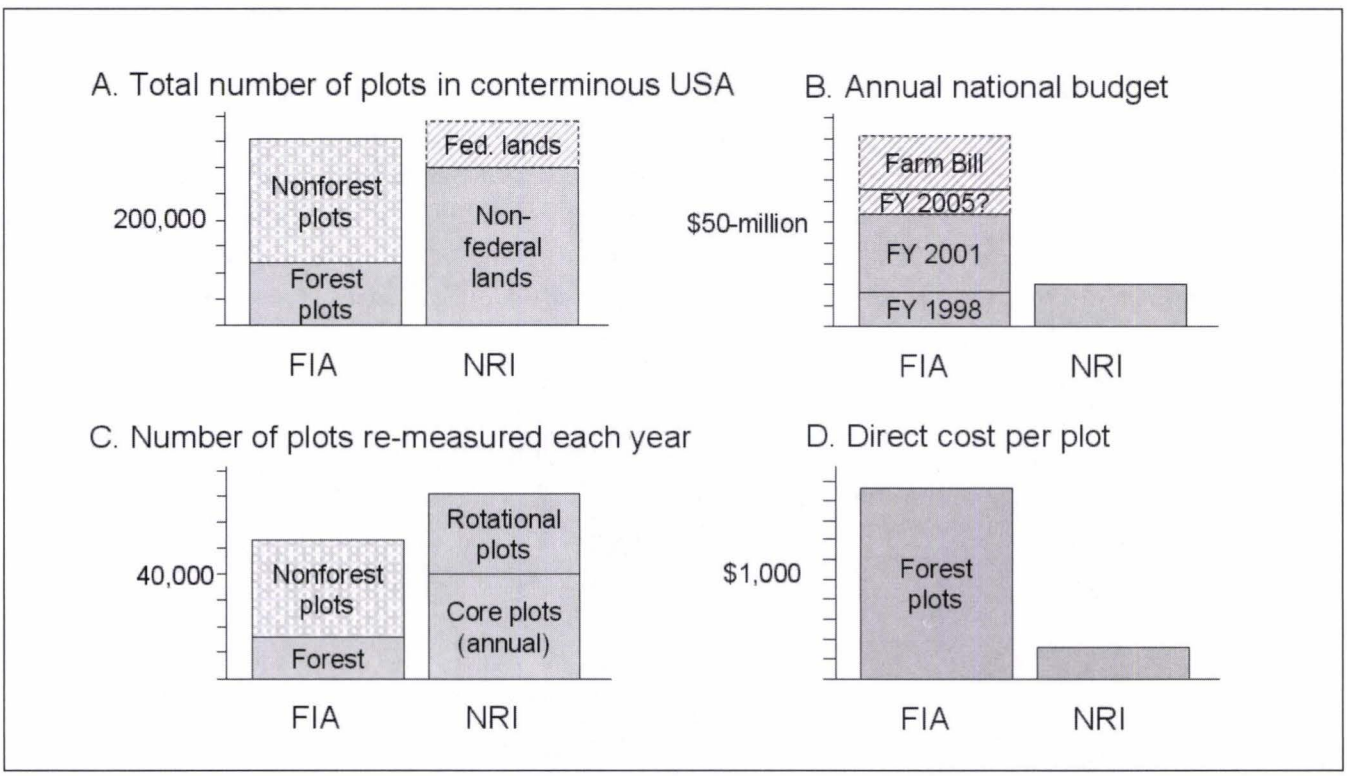
Comparison of FIA and NRI Surveys

FIA maintains one field plot for every 6,000 acres, regardless of land ownership or presence of forest cover. FIA uses a systematic sampling grid and equal selection probabilities for each plot. NRI uses one plot per 8,000 acres of non-Federal lands, with more intensive sampling where land use and resource patterns are more heterogeneous. These unequal selection probabilities increase statistical efficiency and accommodate special analyses. NRI does not currently measure plots on Federal lands.

FIA relies primarily on re-measurement of field plots. While expensive, field measurements are required for accurate estimates of tree- and site-characteristics. However, FIA uses remote sensing to improve precision of statistical estimates[8] of forest area using low-resolution aerial photography or Landsat satellite imagery. NRI primarily uses high-resolution aerial photography to measure status and changes in land cover, land use, and land management practices. These changes are especially important in NRI erosion estimates. NRI uses a limited amount of fieldwork to measure features they believe do not often change over time, or cannot be accurately obtained with aerial photography.

The FIA field plot has four subplots that together encompass about 0.17-acres (fig. 2). The NRI plot, referred to as the Primary Sampling Unit, or PSU, is typically 160-acres (fig. 2). Most NRI plots have three secondary sampling points, at which detailed photointerpreted measurements are made. In recent years, NRI has made these measurements with custom 1:8,000-scale aerial photos. These sample photographs pro-

Figure 3.—*Comparison of FIA and NRI based on number of plots and cost (USDA 1999, 2002).*



[8] FIA uses post-stratification, double sampling for stratification, or double sampling for regression to reduce variance for estimates of forest area. This also reduces variance for estimates of population totals, such as volume.

vide unusually high resolution for interpretation of forest cover and land use. For example, resolution at this scale is sufficient to detect single-family houses under a tree canopy and individual tree mortality.

There are about 360,000 permanent FIA field plots in the U.S. (fig. 3A), located on both private and public land. About 120,000 of those are forested and are intensively measured by field crews. The remaining 240,000 are nonforested and are not measured in significant detail. NRI has about 300,000 NRI plots in the U.S., all of which are measured regardless of their land use. However, NRI does not measure Federal lands; an additional 75,000 NRI plots would be required to cover this land (fig. 3A). Most NRI plots include three secondary sampling points (fig. 2).

Since 1999, both FIA and NRI have adopted different forms of annualized systems for re-measuring permanent plots. The 1998 Farm Bill required FIA to change from re-measuring all FIA plots in an entire State once every 10 to 20 years, to re-measuring 10 percent to 15 percent of all FIA plots in every State every year. FIA plots are separated into five groups, called panels, which are uniformly distributed over the landscape. With current funding, all FIA plots in a single panel are re-measured within a 12- to 24-month period. Then, fieldwork restarts on the next panel. When partial implementation of the 1998 Farm Bill is fully funded, it will take about 7 years to re-measure all FIA plots in the Eastern United States, and about 10 years in the Western U.S. (USDA 2002). On the other hand, NRI plots are divided into two groups: "Core" plots are re-measured every year; NRI "Rotational" plots are re-measured at variable intervals, depending on analysis issues and funding. FIA currently re-measures over 50,000 of its 360,000 field plots each year (USDA 2002); about 19,000 of these field plots contain trees and the remaining 31,000 have no forest cover (fig. 3C). NRI re-measures all 42,000 Core plots and 32,000 of its 258,000 Rotational plots each year (fig. 3C); it acquires and processes over 74,000 aerial photographs (fig. 2) each year.

Detailed tree- and site-conditions on an FIA plot can be accurately measured only in the field. On average, a two-person field crew can re-measure one 0.17-acre forested FIA plot each day. The average cost is $1,800 per plot (fig. 3D), although cost varies by geographic area. NRI statistics are more sensitive to changes in land cover and land use, which can be reliably measured with photointerpretation (fig. 2). The average direct cost for re-measuring a 160-acre NRI plot is about $150 (fig. 3D), of which half is for procurement of the 1:8,000-scale aerial photograph and the remaining half is for labor costs.[9]

Search for Synergy

The experiments reported here evaluated the advantages of fusing the FIA and NRI plot databases with remotely sensed data and statewide GIS database.[10] We hypothesized that this combined database would yield synergies during important analyses. We tested this hypothesis by analyzing land cover and changes in land use with NRI data from 1987 and 1997; FIA data from 1977, 1990, and 1996; and Landsat satellite data from 1986 and 1996.

The most time-consuming portion of these experiments was assembly and harmonization of data from disparate sources. This included combining similar but different FIA and NRI classification systems into a single system. The FIA classification system has detailed categories for stand-level forest conditions but little detail for nonforest conditions, while the NRI system focuses on agricultural uses and land cover on non-Federal lands, with but little detail on forest conditions. A cross-walk was developed that reclassified FIA and NRI categories into five common categories: forest, crops, urban, herbaceous cover, and other land uses (Rack *et al.* 2002). However, some differences between FIA and NRI could not be fully reconciled in the database; these imperfections impact our results to some unknown degree.

FIA and NRI classification systems are based on "land use," which is more difficult to apply with digital classification of satellite data than is "land cover." For example, urban land can include forest, grass, and shrub cover, categories easily confused with those same types of cover in nonurban landscapes. A photointerpreter can use landscape patterns, and the

[9] Personal communication, Dr. J. Jeffery Goebel, Senior Statistician and National Leader for Resource Inventories, Resources Inventory Division, U.S. Department of Agriculture, Natural Resources Conservation Service, Beltsville, MD. Phone: 301–504–2284; fax: 301–504–2230; e-mail: jeff.goebel@usda.gov.

[10] Standard GIS maintained by the State of Minnesota, http://www.dnr.state.mn.us/maps/index.html.

higher resolution available in an aerial photograph, to better deduce land use than can digital classification of satellite data. However, there are inevitable differences among interpreters, and some apparent changes in land cover are likely caused by photointerpretation inconsistencies. Except for classification of forested FIA plots, photointerpretation is used by both FIA and NRI to classify land use.

Spatial Patterns of Land Use Change among FIA and NRI Plots

The first experiment attempted to better understand changes in land use between 1977 and 1997 by analyzing the spatial patterns among changes on sample plots (Rack *et al.* 2002). The union of FIA and NRI plots increased the available observations. Those plots that changed were displayed on a map. Kriging produced no discernible relationship to patterns that were visually apparent in the map display. There were obvious clusters of change: near Minneapolis and St. Paul, where forest and agriculture were changed into urban; along the Mesabi Iron Range, where pits and overburden re-vegetated into forest; and near Park Rapids, where forestlands changed to cropland to serve a food processing plant constructed in the 1980s. However, these changes were previously well known, and no new insights were provided through spatial displays of changed FIA and NRI plots.

There were problems in matching locations of nonforested FIA plots on the aerial photographs used for different surveys; some apparent changes from urban to forest were likely caused by registration errors rather than actual changes in land use. Furthermore, differences between FIA and NRI classification systems for land use and land cover made use of the combined data set difficult. Finally, there were unlikely and unexplainable differences occurring at some county boundaries; these were likely caused by inconsistencies in photointerpretation methods for nonforested FIA plots during the 1977 survey. Therefore, the remaining experiments evaluated FIA and NRI plot-level data in combination with remotely sensed data.

Mapping Changes in Land Use

Several experiments evaluated FIA and NRI plots for mapping changes in land use with Landsat data from 1986 and 1996. The test area included one Landsat scene that covered the Minneapolis/St. Paul area. One experiment used supervised classification, which requires large amounts of training data. The results were disappointing (Rack *et al.* 2002). There were too few FIA and NRI plots that had changed within a single Landsat scene, especially those associated with urban development. Another experiment used unsupervised classification of temporal differences in the Kauth-Thomas transformation, which is more orthodox for digital change detection. The resulting clusters were primarily labeled through image-interpretation; however, FIA and NRI plots provided helpful examples of sites that had changed. The resulting 30-m resolution map of changes in land use is a valuable complement to the traditional FIA and NRI statistics on rates of change. However, map accuracy is unknown because there are no independent reference data available. Rack *et al.* (2002) describe this complex operation in more detail.

Supervised Mapping of Land Cover

Another experiment evaluated FIA and NRI plots as training data for supervised classification of land cover with multiple Landsat scenes for northeastern Minnesota.[11] The remote sensing procedures were designed for the National Land Cover Data (NLCD-2001) Program[12] (Homer *et al.* 2002). NLCD is a consortium of Federal agencies that is building a national Multi-Resolution Land Characterization (MRLC 2001) database of Landsat 7 ETM+ imagery, nominally from the year 2001. The database includes three dates of imagery per Landsat scene: early season, peak greenness, and late season. Radiometric calibration of the Landsat imagery improves consistency of mosaics that include multiple Landsat scenes. NLCD-2001[13] will be a 30-m resolution geospatial database for the entire U.S., including seamless, Web-based delivery of standardized Landsat data (multi-season, Normalized Tasseled Cap transformation); independent ancillary data layers (30-m resolution slope, aspect and elevation); independent, Landsat-

[11] NRLC-2001 Map Zone 41 (http://landcover.usgs.gov/pdf/homer.pdf).

[12] The MRLC and NLCD consortia are led by the USGS EROS Data Center. See www.mrlc.gov.

[13] http://www.mrlc.gov.

based estimates of percent of imperviousness surfaces and tree canopy density; and supervised classification of land cover with these Landsat and ancillary data. FIA and NRI sample plots provided sufficient training data for supervised classification. FIA and NRI plot data helped increase map accuracy (Huang *et al.* 2002) and agreement of the map with FIA and NRI measurement protocols. If FIA and NRI join NLCD-2001, their customers will have a more accurate and user-friendly, nationally consistent, interagency geospatial database for national and regional assessments.

Software for Managing Aerial Photography

An early experiment looked at cooperation in the development of software that benefited all partners (Rack *et al.* 2002). "*Plotview*" is a user-friendly, secure, intranet graphical system that displays FIA and NRI plot locations, associated aerial photography, and proximate data from a statewide GIS system. *Plotview* facilitated use of FIA and NRI plot data and rapid handling of associated aerial photography and GIS data during classification of Landsat imagery. *Plotview* was a useful demonstration that led to similar developments in the FIA and NRI programs.

Future Directions

The experiments described above produced useful results, but they did not achieve any stunning synergies. Several additional experiments are being considered.

The discrepancy between FIA and NRI estimates of total forestland area is a pervasive problem (fig. 1). Many discrepancies are caused by differences between FIA and NRI classification systems for land use. Some of these differences have already been reconciled during the construction of the database described above. Perhaps there are additional ways to better align the classification systems and protocols used by FIA and NRI; separation of classifications systems into land use and land cover holds promise.

Assessments of forest resources with FIA data can benefit from information on soils from NRCS,[14] and assessments of forest soils with NRCS data can benefit from information about forest conditions from FIA. Such assessments could be enhanced by adding the corresponding NRCS code for soil group to each FIA plot. The average characteristics of that soil group could be associated with each FIA plot. This would support analyses, such as those for soil carbon described by Prisley[15] (personal communication). Likewise, FIA attributes, such as tree productivity and biomass density, could be summarized across all FIA plots for each soil group, and those mean FIA values stored in the NRCS national soils database as representative descriptors. Since assembly of disparate databases can be the single largest task in multi-resource assessments, cross-referencing FIA and NRC databases could reduce these costs to external customers. Some soil groups are rare, and association of a plot with a rare soil group could inadvertently compromise the privacy of the landowner. Additional experiments are being considered to test the value of linking certain attributes in the FIA and NRI databases while protecting privacy of landowners.

The cost of implementing the FIA Federal base program mandated by the 1998 Farm Bill, with its requirement for more current data and re-measurement of 20 percent of all plots each year, is estimated at $90 million per year for full implementation (USDA 1999, adjusted to 2002 dollars), or $68 million for partial implementation (USDA 2002, fig. 3B). In response to the FIA Strategic Plan (USDA 1999), the FIA annual budget has nearly tripled, from $18 million in 1997 to $49 million in 2001 (USDA 2002). However, these funds are not yet adequate for even partial implementation of the Federal base program. The current FIA strategy (USDA 1999) transforms traditional FIA periodic surveys into annual surveys by changing the plot re-measurement schedule. Alternatively, combination of FIA and NRI statistical estimates might achieve the 1998 Farm Bill mandate with current FIA funding. The direct cost of re-measuring an NRI plot is about one-tenth the direct cost of re-measuring an FIA plot (fig. 3D) The cost of the current FIA strategy might be reduced if the NRI system could frequently monitor changes in forest area, which can be rapid in many areas during 5 years, and the FIA system could less frequently re-measure tree- and site-conditions within undisturbed forest stands, which usually change more slowly (Smith *et al.* 2001).

[14] http://nasis.nrcs.usda.gov/index.html.

[15] Prisley, Stephen. Personal communication, Virginia Polytechnic Institute and State University, Blacksburg, VA: November 20, 2002.

Assume NRI estimates of forest area could be subdivided into the following categories through photointerpretation: nonstocked stands, clearcuts, partial cuts; seedling/sapling stands, poletimber stands, deciduous sawtimber stands, coniferous sawtimber stands, and mixed sawtimber stands. Further assume that these NRI estimates could be statistically calibrated for photointerpretation errors and differences between the FIA and NRI classification systems. FIA field data could estimate volume per forested acre for each of these stand conditions. The product: (forest acres) x (volume per forested acre) = (total volume). Other FIA estimates of population totals could be similarly estimated. This approach might have little impact on current operations within FIA and NRI, while producing high-quality statistical estimates under current funding levels. Additional experiments will test these assumptions and conjectures in Minnesota.

Summary

Agencies can work together at three levels:[16] coordination, cooperation, and collaboration. Coordination is communication among agencies involved, but each separately conducts its own work. The next higher level is cooperation, which occurs when agencies work together because it would directly benefit each one's mission. Collaboration emerges as agencies work together to develop synergies. While coordination is the easiest to implement, it brings the least benefits. Collaboration takes considerable time, effort, and perseverance, but it can be the most beneficial to participating agencies, their customers, and the public. Future experiments in the integration of FIA and NRI products will examine how to better achieve true collaboration.

Acknowledgment

Funding was provided by the FIA Research Work Unit RMRS-4804 in Fort Collins, CO, and the Washington and Minnesota Offices of the NRCS. The Resource Assessment Unit of the Minnesota Department of Natural Resources, Division of Forestry conducted the experiments described above. The North Central FIA Research Work Unit NC-4801 in St. Paul, the Madison NRI Project Office, and the USGS EROS Data Center provided in-kind support. Jeff Goebel, Collin Homer, Andy Lister, Ron McRoberts, and Jim Westfall provided valuable reviews. Andy Gillespie helped during the early phases of this joint effort.

Literature Cited

Homer, Colin G.; Huang, L. Yang; Wylie, B. 2002. Development of a circa 2000 landcover database for the United States. In: Proceedings, American Society for Photogrammetry and Remote Sensing; 2002 April 19–26; Washington, DC (CDROM).

House, Carol C.; Goebel, J. Jeffery; Schreuder, Hans T.; *et al.*, 1998. Prototyping a vision for inter-agency terrestrial inventory and monitoring: a statistical perspective. Environmental Monitoring and Assessment. 51(1-2): 451–463.

Huang, Chengquan; Yang, Limin; Homer, Collin; *et al.* 2002. Synergistic use of FIA plot data and Landsat 7 ETM+Images for large area forest mapping. In: 3d Annual Forest Inventory and Analysis Symposium, 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC–230. St. Paul, MN; U.S. Department of Agriculture, Forest Service, North Central Research station: 50–55 (http://landcover.usgs.gov/pdf/synergistic.pdf).

Rack, Jim.; Boss, Ken; Aunan, Timothy; *et al.* 2002. Annualized integrated inventory and monitoring initiative: final report. Grand Rapids, MN: State of Minnesota, Department of Natural Resources, Resources Assessment Unit. 55 p.

Smith, W. Brad; Vissage, John S.; Darr, David R.; Sheffield, Raymond M. 2001. Forest resources of the United States, 1997. Gen. Tech. Rep. NC-219. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 190 p.

[16] Dr. Richard Guldin, Director of Science, Policy, Planning, Inventory and Information; U.S. Department of Agriculture, Forest Service, Washington, DC 20090. Phone: 703–605–4177; e-mail: rguldin@fs.fed.us.

USDA. 1992. Report of the Blue Ribbon Panel on forest inventory and analysis. Washington, DC: U.S. Department of Agriculture, Forest Service. 10 p. (http://fia.fs.fed.us/library.htm#Business).

USDA. 1995. A report to the Chief of the Natural Resources Conservation Service by the Blue Ribbon Panel on natural resource inventory and performance measurement. Data rich and information poor. Washington, DC: U.S. Department of Agriculture, Natural Resources Conservation Service.

USDA. 1998a. Report of the Second Blue Ribbon Panel on forest inventory and analysis. Washington, DC: U.S. Department of Agriculture, Forest Service. 17 p. (http://fia.fs.fed.us/library.htm#Business).

USDA. 1998b. Integrating surveys of terrestrial natural resources: the Oregon Demonstration Project. Inv. Monitor. Rep. 2. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Service Inventory and Monitoring Institute. 20 p. (http://www.pwrc.usgs.gov/brd/odp/odp.htm).

USDA. 1999. Strategic plan for forest inventory and monitoring. Washington, DC: U.S. Department of Agriculture, Forest Service. 48 p. (http://fia.fs.fed.us/library.htm#Business).

USDA. 2002. Forest inventory and analysis fiscal year 2001 business report. FS-747. Washington, DC: U.S. Department of Agriculture, Forest Service. 31 p. (http://fia.fs.fed.us/library.htm#Business).

The Effects of Removing Condition Boundaries on FIA Estimates

David Gartner and Gregory Reams[1]

Abstract. —When Forest Inventory and Analysis (FIA) changed to the national standards for the inventory system, plots with multiple condition codes were introduced to the Southern Station's FIA unit. FIA maps up to five different conditions on completely or partially forested 1/24-acre subplots. This change has made producing inventory estimates more complex because the data are analyzed by condition classes (partial plot) rather than on a whole plot. Methods for analyzing by condition classes are less intuitively obvious than methods based on a single condition for an entire plot. We compared the current standard of fully mapped plots to two methods that reduce the number of mapped conditions per plot using the following sets of rules. Rule 1 assigns the predominant condition to the entire plot, including plots that are partially forested. Rule 2 maps a single nonforest condition and a single forest condition, with the single forest condition predominant. The effects of these changes were shown by calculating forest area by forest type group and ownership and volume by species and diameter class. The effect of using just one condition per plot (rule 1) increased the estimated total forested area by 0.4 percent. Using rule 1 decreased the calculated total volume for the state by 3.9 percent. Using only one forested condition per plot (rule 2) did not change the total area estimate or the volume estimates. Both methods decreased the estimated variance for the total volume by 12 to 16 percent. The percent changes in the estimated values were greatest in the least occurring table entries, e.g., the rarest combinations of forest type and ownership group or combinations of species and diameter class.

When Forest Inventory and Analysis (FIA) changed to the national standards for the inventory system, plots with multiple condition codes were introduced to the Southern Station's FIA unit (Reams and Van Deusen 1999). These multiple conditions per plot require determining the boundary between the conditions in the field. In natural stands, forest type boundaries are frequently a continuum, as opposed to abruptly changing from one forest type to another. Therefore, determining the boundary between two forest conditions can be difficult and nonrepeatable. This report addresses two possible changes to the current procedures. One would be to ignore all the boundaries and just use the largest condition on the plot (rule 1). The other possible change would be to ignore boundaries between forested conditions, and use one (the largest) forested condition per plot and one nonforested condition. Our study sought to determine the effects of these possible changes on State estimates of area and volume.

Methods

To simulate the removal of all boundaries, the condition variables, such as forest type and stand age, of the largest condition were assigned to the entire plot. To simulate the possible removal of boundaries between forested conditions, the condition variables of the largest forest were attributed to all the forested area on the plot. For those plots where the two largest forests were the same size, one of the two was chosen at random to be attributed to the larger area, whole plot, or total forested area on the plot. The data came from South Carolina's first panel.

Each plot (sample unit) consists of four 1/24-acre subplots. The condition boundaries actually occur on the subplots. However, there were not sufficient data to analyze the two alternative methods at the subplot level in the database for South Carolina's first panel.

The table values were calculated in the standard manner (FIA Stat Band, in review), except for using double sampling for forested area estimates. First, the phase 1 photo interpretation data were used to estimate the amount of forested land per county. The phase 1 photo interpretation plots are ground truthed on phase 2 plots and a set of ground truth intensification plots. Because condition data are only gathered on the

[1]Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, Knoxville, TN 37919. Phone: 865–862–2066; e-mail: dgartner@fs.fed.us, and Project Leader, U.S. Department of Agriculture, Forest Service, Southern Research Station, Knoxville, TN 37919.

phase 2 plots, the two possible rule changes could not be applied to the photo interpretation intensification plots.

The phase 2 plot data were used to post-stratify the forested area by forest management type. The table values were then calculated using post-stratified sampling estimation techniques (Cochran 1977). To be able to detect any differences between the effect on estimating area and the effect on estimating volume, one area table and one volume table were calculated for each method of handling boundaries. The two tables are 'Table 2—Area of forest land by forest-type group and ownership class' and 'Table 19—Volume of live trees on timberland by species and diameter class' (Thompson 1998).

Specifically, the values for the area table were calculated using equation 1.

$$\hat{Y}_{ot} = A_{total} * \hat{P'}_f * \hat{R}_{ot} \quad (1)$$

where $\hat{Y}_{ot}$ is the estimate of the area for each owner group (o) and forest type (t) combination for a given county, A_{total} is the total land area in the county obtained from the U.S. Census Bureau, $\hat{P'}_f$ is the ratio of forested land to total land area in the county as estimated from the phase 1 data, and $\hat{R}_{ot}$ is the estimated proportion of forest land by ownership group and forest type in the county.

$\hat{P'}_f$ is estimated using double sampling techniques using equation 2 (Reams 2000).

$$\hat{P'}_f = (\hat{P}_f x C_f) + (\hat{P}_n x C_n), \quad (2)$$

where $\hat{P'}_f$ is the proportion of phase 1 plots photo interpreted to be forested, C_f is the proportion of ground truthed plots whose initial photo interpretation calls were forested and were ground truthed as forested, $\hat{P}_n$ is the proportion of phase 1 plots photo interpreted to be nonforested, and C_n is the proportion of ground truthed plots whose initial photo interpretation calls were nonforested that were ground truthed as forested.

$\hat{R}_{ot}$ is the proportion of forested area in a given ownership group (o) and forest type (t) combination as estimated by a ratio of means estimator using equation 3 (Cochran 1977, Zarnoch and Bechtold 2000).

$$\hat{R}_{ot} = \frac{\sum_{i=1}^{n_f} A_{ot_i}}{\sum_{i=1}^{n_f} A_i}, \quad (3)$$

where A_{ot_i} is the area in ownership group by forest type combination on phase 2 plot i, A_i is the amount of forested area on plot i, and n_f is the number of phase 2 plots with at least one forested condition.

Because $\hat{Y}_{ot}$ is a product of two random variables, the variance of $\hat{Y}_{ot}$ (Var($\hat{Y}_{ot}$)) contains the product of the variances of the two random variables and the cross products of those variances and the squared means of the random variables using equation 4 (Goodman 1960, McCollum 2002).

$$Var(\hat{Y}_{ot}) = A_{total}^2 * \{[\hat{P'}_f{}^2 * Var(\hat{R}_{ot})] + [Var(\hat{P'}_f) * \hat{R}_{ot}{}^2] - [Var(\hat{P'}_f) * Var(\hat{R}_{ot})]\} \quad (4)$$

where $\hat{Y}_{ot}$, A_{total}, $\hat{P'}_f$, and $\hat{R}_{ot}$ have the same definitions as equation 1. The variances of $\hat{P'}_f$ and $\hat{R}_{ot}$ are shown in equations 5 and 6.

$$Var(\hat{P'}_f) = \frac{\hat{P}_f * \hat{P}_n}{n} * (C_f - C_n)^2 + [\frac{(\hat{P}_f)^2 * C_f * (1 - C_f)}{m_1}] + [\frac{(\hat{P}_n)^2 * C_n * (1 - C_n)}{m_2}] \quad (5)$$

where $\hat{P'}_f$, $\hat{P}_f$, $\hat{P}_n$, C_f, and C_n have the same definitions as in equation 2, n is the total number of photo interpretation points, m_1 is the number of photo interpretation plots that were ground truthed as forested, and m_2 is the number of photo interpretation plots that were ground truthed as nonforested.

$$Var(\hat{R}_{ot}) = \frac{1}{n_f * (n_f - 1) * [\sum_{i=1}^{n_f} A_i / n_f]^2} * [\sum_{i=1}^{n_f} A_{ot_i}{}^2 + \hat{R}_{ot}{}^2 \sum_{i=1}^{n_f} A_i{}^2 - 2 * \hat{R}_{ot} \sum_{i=1}^{n_f} (A_i * A_{ot_i})] \quad (6)$$

where $\hat{R}_{ot}$, A_{ot_i}, A_i, and n_f have the same definitions as in equation 3.

The values for the volume table for combinations of species group (s) by diameter class (d) are calculated using equation 7.

$$\hat{V}_{sd} = A_{total} * \hat{P'}_f * \sum_{t=1}^{T} \hat{R}_t * \hat{v}_{tsd}, \quad (7)$$

where A_{total} and $\hat{P'}_f$ are as defined in equation 1, $\hat{R}_t$ is the proportion of forested area in a particular forest type, and $\hat{v}_{tsd}$ is the volume per acre for combinations of species groups and

diameter classes. Both $\hat{R}_t$ and $\hat{v}_{tsd}$ are ratios of means estimators of the general form found in equation 3 and variances of the general form found in equation 6. $\hat{R}_t$ is the ratio of the total area in forest type *t* divided by the total forested area summed over plots that contain at least one forested condition. $\hat{v}_{tsd}$ is the ratio of the total volume in the species group by diameter class for forest type *t* divided by the total area in forest type *t* summed over plots that contain at least one condition with that forest type.

The variance of $\hat{V}_{sd}$ is calculated using Goodman's formula for the variance of the product of random variables, as with the variance of $\hat{Y}_{ot}$. However, instead of being a product of two random variables, $\hat{V}_{sd}$ is the product of three random variables. Applying Goodman's formula in succession gives rise to the formula for the variance of $\hat{V}_{sd}$ (equation 8).

(8)

$$Var(\hat{V}_{sd}) =$$
$$A_{total}^{\ 2} * \{\hat{P'}_f^{\ 2} * \sum_{t=1}^{T} Var(\hat{R}_t) * \hat{v}_{tsd}^{\ 2} + Var(\hat{P'}_f) * \sum_{t=1}^{T} \hat{R}_t^{\ 2} * \hat{v}_{tsd}^{\ 2} - Var(\hat{P'}_f) * \sum_{t=1}^{T} Var(\hat{R}_t) * \hat{v}_{tsd}^{\ 2} -$$
$$\hat{P'}_f^{\ 2} * \sum_{t=1}^{T} \hat{R}_t^{\ 2} * Var(\hat{v}_{tsd}) - \hat{P'}_f^{\ 2} * \sum_{t=1}^{T} Var(\hat{R}_t) * Var(\hat{v}_{tsd}) +$$
$$Var(\hat{P'}_f) * \sum_{t=1}^{T} \hat{R}_t^{\ 2} * Var(\hat{v}_{tsd}) + Var(\hat{P'}_f) * \sum_{t=1}^{T} Var(\hat{R}_t) * Var(\hat{v}_{tsd})\}$$

where $\hat{V}_{sd}$ is defined in equation 7, A_{total} and $\hat{P'}_f$ are as defined in equation 1, $\hat{R}_t$ is the proportion of forested area in a particular forest type, and $\hat{v}_{tsd}$ is the volume per acre for combinations of species groups and diameter classes.

If trees in an original forested condition occurred on a plot that was primarily nonforested, using the one condition per plot method removed these trees from the calculations.

Results

Land Area

Because the total forested area is determined by ground truthed photo interpretation points, using rule 2 does not affect the estimates for the total forested area because rule 2 does not change any ground truth calls from forested to nonforested or from nonforested to forested. However, using rule 1 does change some of the ground truth calls for phase 1 photo interpretation points. Using rule 1 increased the estimated total forested area for the State by 0.4 percent and decreased the estimated variance by 0.4 percent. The table cells with the largest area, primarily marginal totals for either forest types or owner groups, had small percent changes, ranging from –6.8 percent to + 2.5 percent. The smaller combinations of forest type and ownership group had larger percent changes, ranging from –100.0 percent to +18.4 percent. The effects of rule 2 on the percent changes followed the same pattern as the effects of rule 1.

Volume

Using rule 1 decreased the estimate for total volume by 3.9 percent and decreased the variance by 15.6 percent. As with the effects on area, the table cells with the greatest amount of volume, primarily the marginal totals for species or diameter classes, show small percent changes, from –6.6 percent to –1.7 percent. The table cells with small amount of volume showed large percent changes, ranging from –100.0 percent to +26.7 percent.

Using rule 2 caused no changes in the estimated volumes from the current method but did change the variance estimates. Using rule 2 decreased the variance of the estimate for total volume by 12.9 percent.

Discussion

Using rule 2 did change the individual cell estimates for area, but not for volume. The reason the volume estimates do not change can be found in equation 7. A close look at the term $\hat{R}_t * \hat{v}_{tsd}$ reveals the cause. In equation 9, the numerator of the first quotient is the area in forest type t summed over all forested conditions, and the denominator of the second quotient is the area in forest type t summed over all forest conditions in forest type t.

$$\sum_{t=1}^{T} \hat{R}_t * \hat{v}_{tsd} = \sum_{t=1}^{T} \frac{\sum_{i=1}^{n_f} A_{ti}}{\sum_{i=1}^{n_f} A_i} * \frac{\sum_{i=1}^{n_t} v_{tsdi}}{\sum_{i=1}^{n_t} A_{ti}} \qquad (9)$$

Since the area in forest type t on plots without any forest type t is 0, these two numbers are equal and will cancel out. Since this is the only condition variable in this equation, the estimates of volume will not depend on how the forest types are determined.

The 12.9-percent decrease in the variance of the estimated total volume for rule 2 is the result of merging forested conditions. When conditions are merged, the resulting volume per acre will be a weighted average of the volume per acre of the conditions that are merged. This weighted average will remove some of the variation in the volume per acre values.

Using rule 1 did affect the estimates of both State totals: an increase in total forest area of less than 0.2 percent and a decrease in total volume of 3.9 percent. When rule 1 is used, some forested and nonforested conditions will be relabeled. The increase in the estimated total forested area suggests that more nonforested area was relabeled as forested than forested area was relabeled as nonforested. When forest is relabeled as nonforest, the trees get removed from the total volume estimate. When nonforest is relabeled as forest, they will contribute to the area estimates, but they don't have any trees to add to the volume estimates, causing a decrease in the volume per acre values for that plot. This decreases the estimated total volume. Changes in the estimates for the individual cells on the reporting tables were much larger, ranging from –100 percent to +30 percent.

Arner (1998) ran a similar study with data from Maine on a different fixed-plot design. His design was a single 1/5-acre plot instead of four 1/24-acre plots. His totals also changed very little and showed larger percent changes in the smaller categories. His relative changes were smaller due in part to the fact that he reported just marginal totals, and in part to the fact the he had more plots. Even though Maine is a smaller state, Arner had more plots because he used the data from a statewide survey, while our study used only one of the panels from South Carolina. Using the rest of the panels from South Carolina would probably lead to smaller percent changes.

In South Carolina, 30 to 35 percent of the plots had more than one condition, but only about 8 percent of the subplots had condition boundaries. Of that 8 percent, about two-thirds were forest-nonforest boundaries. This suggests that using subplot lumping of conditions using either rule 1 or rule 2 methods will also lead to smaller percent changes. Unfortunately, one piece of information required to run this analysis on the subplot level is not in the database for South Carolina's first three panels.

Conclusions

The effect of rule 1 (one condition per plot) on the total area estimate is primarily due to random error. However, rule 1 did bias the estimate for total volume and therefore is unacceptable as an alternative to the current methods. Rule 2 (one forested condition per plot) did not change the estimates for either total forested area or total volume and decreased the variance for the total volume. Rule 2 also did not change the estimates for the volume by species and diameter class combinations. Therefore, rule 2 appears to be an acceptable alternative to the current method.

Recommendations

To determine if the current procedures ought to be changed, we suggest the following steps:

1. To more accurately determine the size of the effects of using rule 2, a similar study should be completed using five sequential panels that include the data required to analyze the effects on the subplot-condition instead of the plot-condition.
2. The results of that study should be discussed with users to determine the effect of the changes in the tables, such as the area by owner group-forest combination table, on the users.
3. The effect of using rule 2 on growth projections should be studied, because some of the condition differences are important to growth projections. For instance, if a 5-inch-diameter loblolly pine stand is growing next to a 9-inch-diameter loblolly pine stand, you would assume that the trees would grow differently in these separate stands than if they were in a single, two-aged stand.

These three steps should allow FIA to make an informed decision about changing to just one forested condition per subplot.

Literature Cited

Arner, Stanford L. 1998. Comparison of a fully mapped plot design to three alternative designs for volume and area estimates using Maine inventory data. Gen. Tech. Rep. NE-243. Upper Darby, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 7 p.

Cochran, William G. 1977. Sampling techniques. 3d ed. New York, NY: John Wiley and Sons. 427 p.

FIA Stat Band (Scott, Charles T.; Bechtold, William A.; Reams, Gregory A.; Smith, William D.; Hansen, Mark H.; Moisen, Gretchen G.). In review. Forest inventory and analysis: national sample design and estimation procedures. Bechtold, William A., and Patterson, Paul L., eds. At http://srsfia1.fia.srs.fe.fed.us/statistics_band/stat_documents.htm.

Goodman, Leo A. 1960. On the exact variance of products. Journal of the American Statistical Association. 55: 708–713.

McCollum, Joseph. 2002. Standard errors in forest area. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; Moser, John W., eds. Proceedings, 3d annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 5–10.

Reams, Gregory. 2000. SAFIS area estimation techniques. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C., eds. Proceedings, 1st annual forest inventory and analysis symposium; 1999 November 2–3; San Antonio, TX. Gen. Tech. Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 32–36.

Reams, Gregory A.; Van Deusen, Paul C. 1999. The southern annual forest inventory system. Journal of Agricultural, Biological, and Environmental Statistics. 4(4): 346–360.

Thompson, Michael. 1998. Forest statistics for Georgia, 1997. Resour. Bull. SRS-36. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 92 p.

Zarnoch, Stan; Bechtold, William. 2000. Estimating mapped-plot forest attributes with ratios of means. Canadian Journal of Forest Research. 30: 688–697.

Comparison of Imputation Procedures for Replacing Denied-access Plots

Susan L. King[1]

Abstract.—In forest inventories, missing plots are caused by hazardous terrain, inaccessible locations, or denied access. Maryland had a large number of denied-access plots in the latest periodic inventory conducted by the Northeastern Forest Inventory and Analysis unit. The denial pattern, which can introduce error into the estimates, was investigated by dropping the 1999 denied-access plots in the 1986 periodic inventory. The denied-access plots represented the population in terms of percentages of forest and nonforest, ownership, land use, and cubic-foot volume. Board-foot volume was less representative. Several single imputation group means—"Euclidean type" distance measures, multiple regression imputation, and listwise deletion with the adjustment of the stratum weights—are compared for estimating the missing cubic- and board-foot volume on forest land. Information on the forested condition of the denied-access plot can be found only through photointerpretation (PI) or satellite imagery such as Multi-Resolution Land Characteristics, (MRLC). Results were inconclusive following an examination of the standard and sampling errors for the state or the root mean square errors for the denied-access plots. As a result, 2 to 12 percent of the data in increments of 2 percent were dropped randomly in a simulation study; the missing plot attributes estimated using each technique. The best simulation study procedure for PI-based forest/nonforest stratification is PI stratum classification. The best simulation study procedure for satellite-based forest/nonforest stratification is the listwise deletion alternative.

In any forest inventory across multiple ownerships, forest crews are occasionally denied access to private and public land. Plots can also be inaccessible due to hazardous terrain and environmental conditions. One solution is to replace the denied or inaccessible plot with a new plot. This raises concerns about sampling strategies that might produce biased results when the ownership of the new plot differs from that of the denied plot, i.e., replacement alters the original sample design. A second solution for denied-access plots is to replace the missing values required for statistical tables with imputed values. This would allow the same plots to be kept over time in the event that permission for access is obtained during the next inventory cycle.

Another concern is the character of the denied-access plots. Do they differ from those in the population? Are there more forested denied-access plots than nonforested and are these more prevalent in a particular ownership? The latter concern is important because different ownerships influence how a forest is managed and grows. Error can be introduced into the estimate when a single group is over-represented. Answering these questions not only provides useful insight to the denied-access pattern, but also is necessary to determine whether the denied-access plots are missing at random. Denied-access plots introduce error into the survey estimates to the extent that they differ from the accessed plots. The reduced sample size also increases the variance of the estimates of the mean from the sample. As the proportion of denied-access plots increases, the bias and variance also can increase leaving the results open to questioning.

Maryland was selected for the study due to its high percentage of denied-access plots, 1.83 percent in the 1999 inventory, a large percentage for the Northeast. Various replacement techniques as well as listwise deletion were applied for cubic- and board-foot volume. Mean volume estimates on forest land and their corresponding standard and sampling errors provided comparison criteria, although estimates of forest land also were examined.

[1] U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA 19073. Phone: 610–557–4048; fax: 610–557–4350; e-mail: sking01@fs.fed.us.

Materials and Methods

Data

Both the 1999 and 1986 inventories of Maryland were used in this study. The 1999 inventory was used to identify which denied-access plots to drop when reprocessing the 1986 inventory. The 1986 inventory provided truth and was used to evaluate the imputation procedures and the nonresponse pattern.

To reduce the sampling error, the State population was divided into two pools: forested and nonforested. Since the denied-access plots were assumed to not have been visited, the classes of these plots are unknown. However, through photointerpretation (PI) or satellite imagery, a forest/nonforest classification is made. Satellite imagery will be used exclusively in the future for the Phase I stratification, but PI provides a comparison. Some of the difficulties encountered in using old data (1986 inventory) included having only the coordinate locations for 948 of the original 1,177 plots, having PI information for only 1,106 plots, and having only Landsat 7 satellite bands 2, 4, and 7.

The satellite imagery selected was a forest/nonforest map acquired from a National Land Cover data set (formerly Multi-Resolution Land Characteristics (MRLC)). This vegetation map was made by the U.S. Geological Survey EROS Data Center (Vogelman *et al.* 2001) and is based on 1992 Landsat 7 Thematic Mapper data and various other intermediate-scale spatial data were used as ancillary data.

For the forest/nonforest call, the MRLC was reclassified so that the forest classes as well as woody wetland received a value of 1 and other pixels received a value of 0. The plots were overlaid on the reclassified image to obtain a forest/nonforest call. The call for PI was based on the land-use classification using photos. For these data, PI had an overall accuracy of 94.4 percent compared to field plots, whereas the satellite forest/nonforest classification was only 80.17 percent as accurate as field plots. These percentages are similar to other data sets and indicate that PI is a more accurate forest/nonforest classifier.

A "5x5 sum" filter was passed over the reclassified forest/nonforest MRLC image. In the resulting image, each pixel represented the count of forested pixels inside a 5x5 window. The plots were overlaid on the filtered image to obtain a forested pixel call for each plot. The forested pixel count ranged from 0 to 25 and was divided into four classes (MRLC5 strata) based on a study using Connecticut data (Hoppus *et al.* 2001). The PI strata are the PI classifications.

One of the imputation procedures substitutes the plot with the nearest Euclidean or spectral distance. For the spectral data, plots were overlaid on bands 2, 4, and 7 from a satellite image created from early 1990s imagery that was composited and radiance- and terrain-corrected by the Earthsat Corporation. Earthsat distributed a band 7-4-2 subset of this original image via the National Aeronautics and Space Administration. The image was from Landsat 7 and had a resolution of 28.5 m. No filter was applied to this image because the original image had been geocorrected to an extremely high positional accuracy.

Imputation Techniques

1. Listwise Deletion: The 21 denied-access plots were dropped and the stratum weights adjusted to reflect the reduced sample sizes in affected strata. The sample size is reduced by the number of denied-access plots and the stratum weights are adjusted to reflect the reduced number of plots in the affected strata.
2. Group Mean Replacement: The denied-access plots were replaced with the State, County, PI, or MRLC5 stratum average or the stratum classification average. The denied-access plots were replaced with the group mean average for cubic- and board-foot volume from their respective forest/nonforest pool. The plot was assigned the forest/nonforest pool classification for both PI and satellite imagery. The sample size was not altered.
3. Replaced with Euclidean (PI) or spectral distance (satellite): On-the-ground classification for forest/nonforest, timberland, cubic-foot volume, and board-foot volume from the nearest plot was substituted for the missing values on the denied-access plot. Again, the sample size was not altered.
4. Statistical Multiple Imputation: The data have a monotone missing pattern because all the plot attributes are either present or missing. Consequently, either a parametric regression procedure, which assumes normality, or a nonparametric method, which uses propensity scores, is appropriate (Rubin 1987, SAS Institute Inc. 2001). The regression approach produced lower sampling errors, so the results for this procedure are reported. As with the group

mean replacement, the plot was assigned the forest and timberland value of the pool. The estimates were imputed separately both by forest/nonforest and by cubic- and board-foot volume. This allows maximum and minimum parameters to be set more accurately in PROC MI (SAS Institute Inc. 2001). The PI classes and the three satellite bands also were used as imputation variables. The previous methods are single-imputation procedures. In multiple imputation, the missing data are filled in *m* times, generating *m* complete data sets. The *m* complete data sets are evaluated by using standard statistical analysis and the *m* complete data sets are combined to produce a final result.

Double Sampling for Stratification

The sampling design used in 1986 was double sampling for stratification. There were four stratum classes in Maryland: forested and nonforested plots inventoried in 1976, and first-time inventoried forested plots and nonforested plots in the 1986 inventory. Each class had 18 land areas consisting of a county or a grouping of adjacent counties. Thus, there were 36 forested and 36 nonforested strata. The variable of interest is total or mean cubic-foot volume, board-foot volume, or forestland area. The variance equation was equation 12.12 (from Cochran 1963).

Multiple Imputation Equations

After *m* multiple imputations or *k* simulated complete data sets were generated and each was analyzed as if it were a complete data set, means and variances were combined to form a single estimate. The means can be averaged to form a single estimate but the variances must account for both the within-imputation variance (average of the complete data estimates) and between-imputation variance (see Allison 2002, Gartner and Reams 2002, Reams and McCollum 1998, and SAS Institute Inc. 2001).

Results

Similarity

One goal was to investigate whether denied-access plots differ from those in the general population. Of 1,177 plots in the 1986 inventory, 21 were denied-access plots. The population is 59.98 percent forested and 66.67 percent (14 plots) of the denied-access plots were forested. The denied-access plots are distributed across Maryland rather than a particular region of the State (fig. 1).

As mentioned earlier, different ownerships influence how a forest is managed and grows so the estimates can be biased when one group is overrepresented. With only 14 forested denied-access plots and more than three categories, the minimum cell count of five observations is not met and a Chi-square goodness-of-fit test cannot be performed. Not all the ownership categories are represented in the denied-access plots (fig. 2a). The largest category, Other Private Individuals, is 61 percent of the population and 57 percent of the denied-access sample; class percentages match closely given that there are only 14 plots.

Not all forest types are represented in the denied-access plots (fig. 2b), but the percentages are nearly identical because, again, there are only 14 denied-access plots. For both cubic-foot (fig. 2c) and board-foot (fig. 2d) volume, the extreme classes are not represented in the denied-access plots. The percentages of denied-access plots are more similar to those of the sample population for cubic-foot than for board-foot volume. Consequently, greater bias would be expected in the estimates.

Actual Results for Denied-access Plots

Table 1 shows the State results for PI and MRLC satellite imagery. Under the column for forest land, zero indicates the

Figure 1.—*This map shows both the 21 denied- and all nondenied-access plots in Maryland. The background is a forest/nonforest map and the denied-access plots are in all parts of the State on forest and nonforest land. The western portion of the State is more mountainous, thus it has more plots and more denied-access plots than the other regions.*

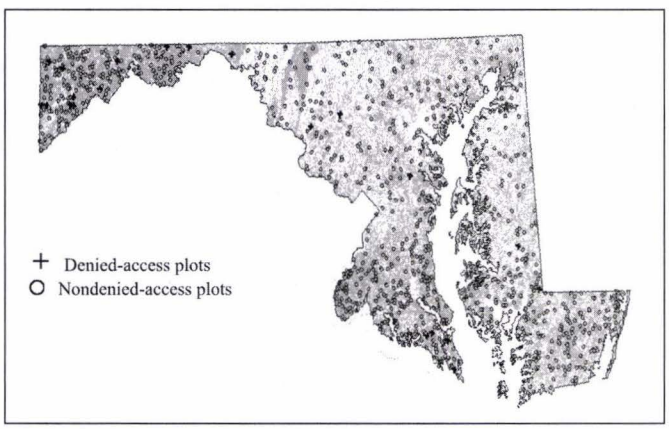


Figure 2.—*Comparison of sampled and denied-access populations by (a) ownership, (b) forest type, (c) cubic-foot volume, and (d) board-foot volume. The sampled population includes both denied and nondenied-access populations. The sampled and denied-access populations are less similar for board-foot volume.*

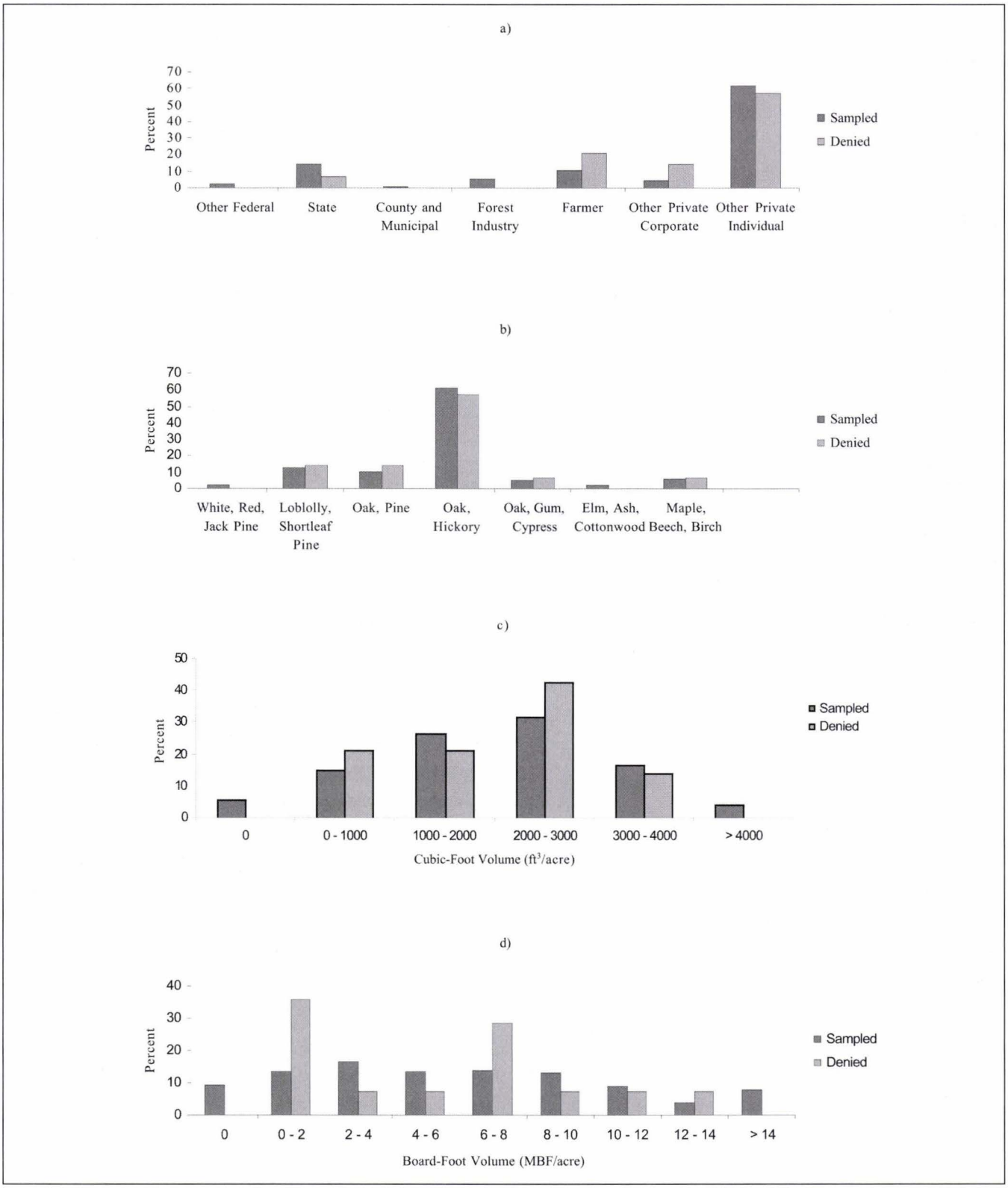


Table 1.—*State results for PI and MRLC satellite*

Procedure		PI cubic-foot volume			Satellite cubic-foot volume			PI board-foot volume			Satellite board-foot volume		
	Forest land	Mean volume	Standard error mean volume	Sampling error mean volume	Mean volume	Standard error mean volume	Sampling error mean volume	Mean volume	Standard error mean volume	Sampling error mean volume	Mean volume	Standard error mean volume	Sampling error mean volume
Complete data	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	727.60	20.40	2.80	727.60	20.40	2.80	2181.68	82.80	3.80	2181.68	82.80	3.80
Listwise deletion	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	726.62	20.47	2.82	726.62	20.47	2.82	2177.17	83.04	3.81	2177.17	83.04	3.81
State average	0	0.07	0.04	55.19	1.25	0.64	50.7	20.21	0.12	55.19	3.69	1.87	50.72
	1	728.08	20.41	2.80	725.26	20.43	2.82	2183.19	82.72	3.79	2173.64	82.72	3.81
County average	0	0.01	0.01	67.11	1.23	0.76	61.76	0.05	0.031	67.11	3.40	2.18	64.24
	1	727.54	20.42	2.81	726.03	20.43	2.81	2183.68	82.72	3.79	2176.38	82.70	3.80
PI or MRLC5 stratum average	0	0.05	0.04	86.65	1.44	0.67	46.53	0.15	0.13	86.65	4.23	1.96	46.40
	1	726.26	20.37	2.80	726.53	20.43	2.81	2175.72	82.62	3.80	2176.75	82.72	3.80
Stratum class-ification average	0	0.04	0.02	67.11	1.65	1.52	91.89	0.12	0.08	67.11	5.26	4.83	91.75
	1	727.68	20.42	2.81	726.37	20.43	2.81	2179.74	82.69	3.79	2177.31	82.71	3.80
Euclidean or spectral distance	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	727.60	20.40	2.80	727.79	20.49	2.82	2197.18	83.56	3.80	2180.16	82.91	3.80
Regression imputatioin	0	0.02	0.13	81.68	0.17	0.17	102.92	0.37	0.28	76.43	0.35	0.42	120.04
	1	729.84	20.54	2.82	726.38	20.54	2.83	2194.33	83.50	3.81	2184.45	82.94	3.80

results for nonforest land and one indicates results for forest land. Euclidean or spectral distances replace the denied-access plot with attributes from the nearest plot, increasing the likelihood that the forest/nonforest classification is correct. A correct classification would result in zero cubic- and board-foot volumes on nonforested land. The standard and sampling errors are very close for both PI and MRLC satellite imagery. Except for the Euclidean distance and regression imputation for estimating mean board-foot volume, mean volumes are close to the mean volume for the complete data. There is no best procedure. In the next step to determine the best procedure, root-mean square errors (RMSE) were calculated for the 21 denied-access plots by procedure. From table 2, the PI stratum average is the best procedure for both cubic- and board-foot volume; there is no superior procedure for the satellite imagery. The MRLC5 stratum average has the smallest RMSE, but the State, county, and stratum classification averages are close. Euclidean or spectral distance and regression imputation had the highest RMSE.

Results of Simulation Study

The next step was a simulation study to determine a winning procedure at the 2-percent denial rate and whether that best procedure might be different at higher rates. The denials were determined randomly without considering maintaining a forest-to-nonforest ratio of 60:40. Only plots with both coordinates and satellite bands 2, 4, and 7 were used so that all procedures at the same denial percentage eliminated the same plots. A different random number seed was used for each percentage for 100 simulations. The squared, absolute, and standard errors of the forested estimates were calculated for each procedure and then averaged by the stratum classification. For each stratum, the averaged results were sorted in ascending order by procedure. The procedure with the smallest comparison statistic was assigned as the best procedure for that stratum. The tabulated RMSE results for the PI and satellite-based simulation at different deletion percentages are shown in tables 3 and 4. For PI, replacement by the PI stratum average was the best procedure. For the satellite-based simulation, listwise deletion was the

Table 2.—*RMSE for 21 denied-access*

Plots	Cubic-foot volume		Board-foot volume	
Procedure	**PI**	**Satellite**	**PI**	**Satellite**
State average	931.46	1054.73	3460.85	3845.50
County average	909.32	1052.48	3455.14	3835.21
PI or MRLC5 stratum average	609.32	1033.88	2745.57	3753.63
Stratum classification average	918.85	1065.12	3495.95	3948.84
Euclidean or spectral distance	1326.85	1241.91	5918.58	4479.65
Regression imputation	1056.59	1100.59	4149.88	3974.99

Table 3.—*Number of times the RMSE for the PI-based simulation was best at different deletion percentages*

	Cubic-foot volume						Board-foot volume					
Procedure	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Listwise deletion	8	6	7	3	8	7	6	6	8	4	6	6
State average	3	7	6	6	5	9	2	6	6	7	5	6
County average	2	3	2	4	3	1	4	5	4	3	3	3
PI stratum average	13	13	16	16	13	16	15	12	12	13	16	16
Stratum classification average	4	4	3	2	1	2	4	2	2	2	2	2
Euclidean Distance	4	3	2	2	3	1	3	3	2	1	1	1
Regression Imputation	2	0	0	3	3	0	2	2	2	6	3	2

Table 4.— *Number of times the RMSE for the satellite-based simulation was best at different deletion*

	Cubic-foot volume						Board-foot volume					
Procedure	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Listwise deletion	20	21	22	21	21	23	20	20	23	24	24	21
State average	0	0	0	1	2	0	2	1	1	2	1	2
County average	6	4	5	5	4	4	5	6	4	2	3	3
PI stratum average	6	7	7	6	9	8	8	6	5	5	7	10
Stratum classification average	0	1	1	1	0	1	1	1	2	2	0	0
Euclidean Distance	1	0	0	0	0	0	0	0	0	0	0	0
Regression Imputation	3	3	1	2	0	0	0	2	1	1	1	0

best procedure with the lowest RMSE in more than 20 of the 36 strata. The best procedures were the same for the other two comparison statistics: absolute and standard errors.

Another comparison entails tabulating by imputation procedure the number of estimates for mean volume (of 3,600) that exceed the 95-percent confidence interval of the mean volume of the total sample estimate. The results for PI and satellite-based forest/nonforest classification are shown in tables 5 and 6. The best procedure has the lowest number of entries. For PI, the best procedure is the PI strata average. However, listwise deletion gave the same results for board-foot volume. Listwise deletion is the best procedure for satellite imagery.

Conclusion

The denied-access population in Maryland was similar to the sample population in terms of the amount of forested land,

Table 5.— *Number of times an estimate from the simulation fell outside the 95-percent confidence interval for the PI-based simulation at different deletion percentages; 180 estimates would be expected to fall outside the confidence interval*

	Cubic-foot volume						Board-foot volume					
Procedure	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Listwise deletion	0	0	1	0	3	0	0	0	0	0	1	0
State average	0	0	0	1	0	4	0	0	0	1	1	1
County average	0	0	0	1	1	3	0	0	0	1	0	0
PI stratum average	0	0	0	0	0	2	0	0	0	0	1	0
Stratum classification average	0	0	0	2	0	3	0	0	0	1	1	0
Euclidean Distance	1	5	11	12	8	14	0	1	7	6	10	17
Regression Imputation	0	1	1	3	1	6	0	1	0	5	3	4

Table 6.— *Number of times an estimate from the simulation fell outside the 95-percent confidence interval for the satellite-based simulation at different deletion percentages; 180 estimates would be expected to fall outside the confidence interval*

	Cubic-foot volume						Board-foot volume					
Procedure	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**
Listwise deletion	0	0	1	0	3	0	0	0	0	0	1	0
State average	1	6	7	16	28	36	0	0	1	4	8	6
County average	2	4	16	29	36	38	1	2	6	15	15	20
PI stratum average	0	1	1	7	10	15	0	0	0	1	3	2
Stratum classification average	0	2	3	16	14	15	1	5	5	13	17	15
Euclidean Distance	22	46	61	104	124	164	20	37	53	82	109	131
Regression Imputation	3	13	26	41	44	51	2	6	11	21	29	40

geography, ownership, forest type, and cubic-foot volume. The two groups were less similar for board-foot volume. Results for estimating forest land are not presented, but they were the least biased since the denied-access forested vs. nonforested land closely matched the sample population.

The State results are inconclusive and the results must be analyzed at the stratum level or by examining the denied-access plots to differentiate between procedures. PI-based forest/nonforest classification is superior to satellite-based classification as evidenced by lower standard and sampling errors as well as lower RMSE. This is not surprising since PI-based forest/nonforest classification more nearly agrees with plot-based estimates than satellite-based classification.

Regression imputation and Euclidean or spectral distance had the poorest performance. The effectiveness of regression imputation depends on the relationship between the available independent variables and the variable to be imputed. The PI classes and satellite bands did not capture the relationship with either cubic- or board-foot volume. The plots are 2.7 km apart and land use and history between nearest neighbor plots can vary greatly. As a result, the Euclidean distance technique was ineffective. Surprisingly, the spectral distance also failed to capture the relationship with cubic- or board-foot volume. Also, the best procedure did not vary by deletion percentages but depended on whether PI or satellite-based classification was used to classify the forested/nonforested condition of the plots. For PI-based stratification, the winning procedure is the group mean replacement, PI stratum classification, whereas for satellite-based stratification, the winning procedure is the listwise deletion alternative with the adjustment of the stratum weights.

Literature Cited

Allison, Paul D. 2002. Missing data. Thousand Oaks, CA: Sage University Papers on Quantitative Applications in the Social Sciences, 07-136. Sage Publications. 93 p.

Cochran, William G. 1963. Sampling techniques. 2d ed. New York, NY: John Wiley and Sons. 428 p.

Gartner, D.; Reams, G.A. 2002. A comparison of several techniques for estimating the average volume per acre for multipanel data with missing panels. In: Reams, Gregory A.; McRoberts, Ronald E.; Van Deusen, Paul C., eds. Proceedings, 2d annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 76–81.

Hoppus, M.; Arner, S.; Lister, A. 2002. Stratifying FIA ground plots using a 3-year old MRLC forest cover map and current TM delivered variables selected by "decision tree" classification. In: Reams, Gregory A.; McRoberts, Ronald E.; Van Deusen, Paul C., eds. Proceedings, 2d annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 19–24.

Reams, G.A.; McCollum, J.M. 1998. The use of multiple imputation in the Southern annual forest inventory system. In: Hansen, Mark; Burk, Tom, eds. Integrated tools for natural resources inventories in the 21st century; 1998 August 16–20; Boise, ID. Gen. Tech. Rep. NC-212. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 228–233.

Rubin, D.B. 1987. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley and Sons. 258 p.

SAS Institute Inc. 2001. SAS/STAT® software: changes and enhancements, Release 8.2. Cary, NC: SAS Institute, Inc. 324 p.

Vogelman, J.E.; Howard, S.M.; Yang, L.; *et.al.* 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing. 67(3): 650–661.

Strategies for Preserving Owner Privacy in the National Information Management System of the USDA Forest Service's Forest Inventory and Analysis Unit

Andrew Lister, Charles Scott, Susan King, Michael Hoppus, Brett Butler, and Douglas Griffith[1]

Abstract.—The Food Security Act of 1985 prohibits the disclosure of any information collected by the USDA Forest Service's FIA program that would link individual landowners to inventory plot information. To address this, we developed a technique based on a "swapping" procedure in which plots with similar characteristics are exchanged, and on a "fuzzing" procedure in which the geographic locations of the plots are randomly perturbed by 805 m. A simulation experiment was performed to assess the effects of fuzzing and swapping. Our results indicate the procedures can provide meaningful information and comply with the law. Further refinements of the technique are ongoing.

The USDA Forest Service's Forest Inventory and Analysis (FIA) program is responsible for conducting a national forest inventory (Gillespie 1999). FIA uses a network of tens of thousands of ground plots to collect information on the quantity, quality, composition, location, and other characteristics of the forests and on land ownership and use on these plots. Many of the forested plots are or soon will be georeferencedusing a global positioning system (GPS), making the data in the FIA database appealing to land managers and scientists interested in using FIA data in a spatial context. Historically, FIA did not divulge exact locations of ground plots to protect landowner privacy and to protect the integrity of the sample. FIA attempted to accommodate data consumers and adhere to the existing security policies by performing in-house analyses.

In 2000, the Department of the Interior and Related Agencies Appropriations Act (H.R. 3423) amended the Food Security Act of 1985 (H.R. 2100) to include FIA in a list of activities that may not make data available to the public if the owner of the land on which the data were collected can be identified. Since the FIA data are referenced with GPS, and ownership maps are freely available to the public in county tax offices, making public the plot data with GPS or digitized location is tantamount to revealing the owner's name and thus violating the law.

In addition to addressing legal concerns, maintaining privacy of the plot locations is essential to FIA's mission. If the plot location were freely available, individuals could either intentionally or unintentionally alter the ecological conditions on the plot, impacting the integrity of data that are collected the next time the plot is measured (in 5 to 10 years). Furthermore, the value of the data is degraded if it is felt that land managers might intentionally alter land management around FIA plots to affect (or avoid affecting) the survey.

Nonetheless, FIA wants to assist users in utilizing the spatial nature of the FIA data while preserving privacy. To reach this goal, we developed a technique whereby the plot coordinate data are slightly altered (fuzzed) and some of the plot data are exchanged (swapped). The purpose is to maintain the functional value, or "ecological signal" of the data while introducing enough uncertainty to decouple the plot-landowner relationship. We then tested the effects of this fuzzing and swapping on the calculation of average board-foot volume within circles of various sizes. The goal of the experiment was to characterize the distribution of errors that data consumers might get when using fuzzed and swapped data.

Methods

The geographic location data collected on 2,037 plots in Maine between 1999 and 2001 were fuzzed using ArcView GIS software (ESRI, Redlands, CA 92373) such that each "new" plot

[1]Forester, Project Leader, Operations Research Analyst, Research Forester, Supervisory Forester, and Forester, respectively, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA 19073. Phone: 610–557–4038; e-mail: alister@fs.fed.us.

site was located on land within the same county in a random direction by up to 805 m from its original location.

To perform the swap, forested plots on private land were placed into groups based on ownership: forest industry, nonindustrial corporate, other nonindustrial private, and nonindustrial individual. If there were not at least three unique owners within each group within a county, the groups were combined as follows: forest industry with nonindustrial corporate, and other nonindustrial private with nonindustrial individual. If there were still fewer than three owners, all categories were combined into a single "private lands" category. If there were not at least three owners in this private lands group, adjacent counties were combined until all criteria were met.

From within these groups, 12.5 percent of the plots were chosen for exchange with ecologically similar plots within the same group to produce a 25-percent swap. The Euclidean distance-based similarity measure was calculated using the following equation:

$$\text{Similarity Value} = (\text{northing}_a - \text{northing}_b)^2 + (\text{easting}_a - \text{easting}_b)^2 + (\text{forest type group}_a - \text{forest type group}_b)^2 + (\text{productivity class}_a - \text{productivity class}_b)^2$$

where *a* is a plot in the group selected for exchange, and *b* is a plot in the group not in the original selection but still in the same group and county.

Smaller values indicate more similarity. The similarity-defining variables for swapping were chosen because they are static; it would be undesirable to swap plots based on characteristics that would likely change between inventories. Northing and easting are Albers Equal Area coordinates in meters, forest type group is an FIA tree species group identification number that ranges from mostly conifers at the low end to mostly deciduous species at the high end, and productivity class is a value measured in the field by FIA crews that is based on site index, which is the relationship between a representative tree's height and its age.[2]

To test the results of this procedure, a simulation experiment was performed. ArcView GIS software was used to create 1,000 randomly located circles with radii of 5, 10, and 20 km in Maine. Within each circle, the average board-foot volume (bfv) of the unperturbed plots and that of the fuzzed and swapped plots was calculated by summing the bfv of each tree located on each plot within the circle. For each circle, the absolute difference (AD) between pre- and post-fuzzed and swapped bfv averages was calculated, and histograms of ADs were constructed. Scatterplots were created to visually assess the relationship between unperturbed averages and fuzzed and swapped averages, and simple linear regressions were calculated for the data in these scatterplots to describe the relationships.

Results and Discussion

The histograms of ADs from the 5-, 10-, and 20-km circles are shown in figure 1. The means and coefficients of variation of the ADs (in parentheses) for the 5-km, 10-km, and 20-km circles are, respectively, 877.3 (179 percent), 478.8 (102 percent), and 207.0 (86 percent). The degree of skewness (Υ) decreased with increasing circle radius (Υ_5=3.2, Υ_{10}=1.4, Υ_{20}=1.3).

There were about 30 plots in the 20-km circles, 8 in the 10-km circles, and 2 in the 5-km circles. The 5-km circles have the largest percentage of ADs in the lowest error category and the largest range, followed by the 10-km and then the 20-km circles. This is because, for smaller circles, the bfv averages in the tails of the unperturbed data's distribution are very susceptible to either no change or extreme change after fuzzing and swapping, leading to either very small or very large ADs. For the larger circles, however, swapping more likely will occur from within a circle than from without, and the smaller perimeter/area ratio lowers the chances of plots being fuzzed into or out of a circle.

The scatterplots of the unperturbed versus the fuzzed and swapped bfv averages are shown in figure 2. The confidence intervals for the parameters of the regression lines whose equations are shown on figure 2 are shown in table 1. The slopes and intercepts of all regression lines (fig. 2) indicate that the bfv averages calculated from fuzzed and swapped plots are overestimated for low bfv averages and underestimated for high bfv averages. The y intercepts of the regression lines

[2] USDA Forest Service. 2000. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.4. USDA Forest Service, internal report. On file at USDA Forest Service, Washington Office, Forest Inventory and Analysis, Washington, DC.

Figure 1.—*Histograms of absolute differences in board-foot volume (bfv) (AD=abs(unperturbed average bfv – fuzzed and swapped bfv)) obtained within circles of 20-, 10-, and 5-km radius. These graphs represent the effects that the fuzzing and swapping procedure had on data retrievals. N=1000.*

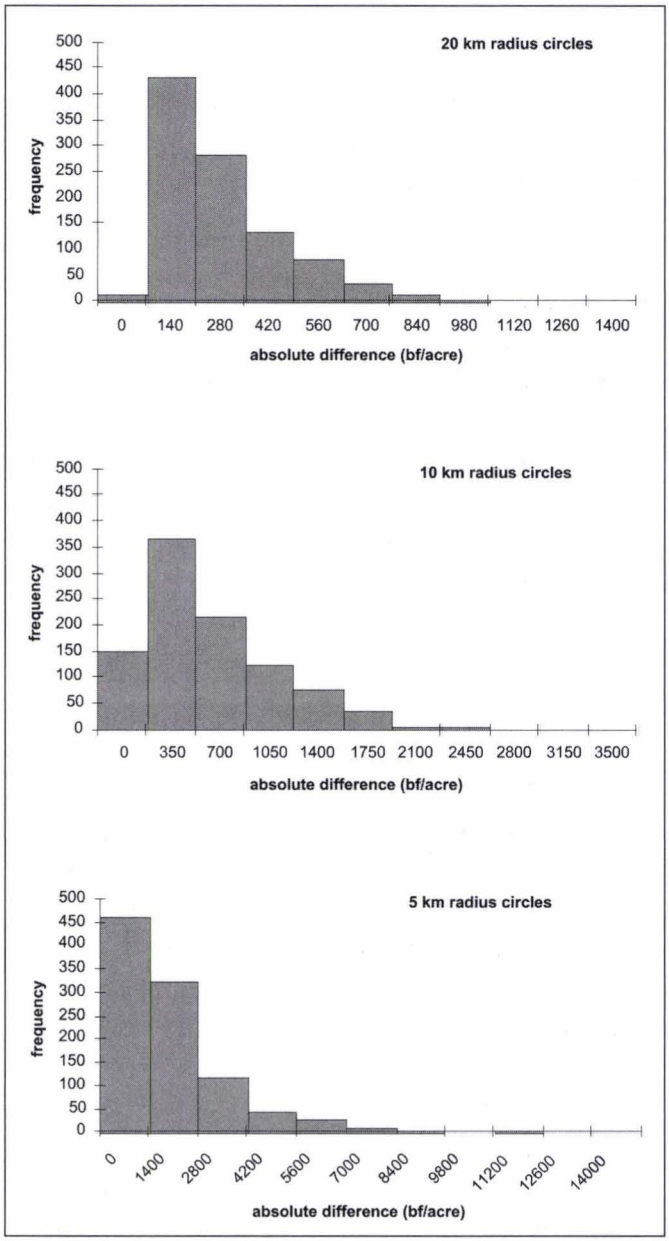


Figure 2.—*Scatterplots and simple linear regression lines and equations of fuzzed and swapped bfv averages versus unperturbed bfv averages within circular areas of various radii. N=1000.*

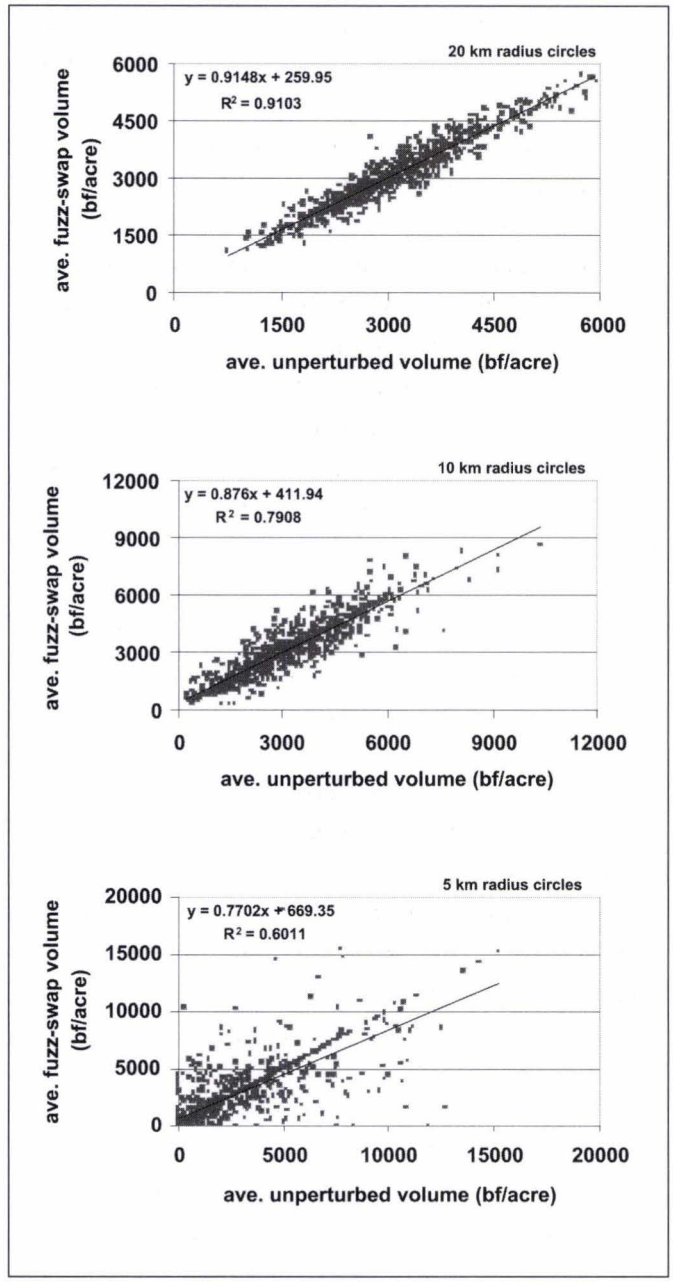

20 km radius circles
y = 0.9148x + 259.95
R² = 0.9103
ave. fuzz-swap volume (bf/acre)
ave. unperturbed volume (bf/acre)
10 km radius circles
y = 0.876x + 411.94
R² = 0.7908
ave. fuzz-swap volume (bf/acre)
ave. unperturbed volume (bf/acre)
5 km radius circles
y = 0.7702x + 669.35
R² = 0.6011
ave. fuzz-swap volume (bf/acre)
ave. unperturbed volume (bf/acre)

Table 1.—*Ninety-five-percent confidence intervals of the coefficients of the simple linear regression lines that describe the scatterplots of unperturbed vs. fuzzed-swapped plot data for three circle radii*

	5 kilometer			10 kilometer			20 kilometer		
	coeff	lower	upper	coeff	lower	upper	coeff	lower	upper
Slope	0.77	0.73	0.81	0.88	0.85	0.90	0.91	0.90	0.93
Intercept	669.4	514.2	824.5	412.0	314.0	509.8	260.0	202.4	317.5

decrease with increasing circle radius (fig. 2); none of the 95-percent confidence intervals for the intercept contain 0 (table 1). The smaller circles have a larger range of ADs (fig. 1); this forces the y intercept of the regression lines away from zero for the smaller circles. It is noteworthy that there are no zero values found in the scatterplots for the 10- or 20-km circles. There are, however, 1,000 points defining the trajectory of the regression line, making us feel comfortable with our use of the y intercept as a statistic that describes these scatterplots.

All of the 95-percent confidence intervals for the slopes of the regression lines fall below 1 (table 1). The slopes and R^2 values for the regression lines approach 1 with increasing circle radius (fig. 2). This is because, for larger circles, the averaging effects of the fuzzing and swapping tend to act uniformly throughout the entire distribution of values, lowering the variance of the ADs and the deviation of the regression line from a slope of 1.

Across all circle sizes, the bfv averages in the tails of the distribution will always tend toward the sample mean after fuzzing and swapping occurs. In general, the bfv average of a plot moving into a circle with a bfv average in one of the tails of the distribution will be nearer to the sample mean value than it will be to that of the other plots in that circle.

Our results might be difficult to generalize. For example, there is no guarantee that other plot attributes will have the same or less variability than average bfv. Larger data retrievals will be less subject to large fluctuations in summary values than will small ones because a smaller percentage of the total number of plots will be affected by fuzzing. Likewise, data retrievals within different shaped areas might affect summaries more than circular retrievals due to the effects of the geometric complexity of landscape patterns. However, we see no reason to believe that the same principles that governed our current results will not hold with other variables.

In conclusion, the fuzzing and swapping technique outlined here shows great promise. It was conceived as a way to provide useful data to interested parties outside of FIA without violating the law, compromising the ecological integrity of the plots, or introducing concerns about treatment bias. An effort to maintain the functional value of the data is made by conducting geographic fuzzing within a short distance and by swapping plots with similar ecological conditions. The results of our simulation experiment suggest that for average bfv within user-defined circular areas of various sizes, the functional value of the data is kept relatively high, i.e., the fuzzing and swapping technique does not change the fundamental quality of the data dramatically. The functional value is highest for retrievals containing more plots. The fuzzed and swapped data will be most useful for consumers interested in creating summaries over large areas, or for those interested in producing their own coarse-scale graphical representations of the occurrence of FIA-measured attributes. Furthermore, correlative studies with other spatial layers also can be conducted as long as the analyst understands the impact of the slight loss of the data's functional value.

Literature Cited

Gillespie, A.J. 1999. Rationale for a national annual forest inventory program. Journal of Forestry. 97(12): 16–20.

COLE: A Web-based Tool for Interfacing with Forest Inventory Data

Patrick Proctor[1], Linda S. Heath[2], Paul C. Van Deusen[1], Jeffrey H. Gove[2], and James E. Smith[2]

Abstract.—We are developing an online computer program to provide forest carbon related estimates for the conterminous United States (COLE). Version 1.0 of the program features carbon estimates based on data from the USDA Forest Service Eastwide Forest Inventory database. The program allows the user to designate an area of interest, and currently provides area, growing-stock volume, and carbon pool estimates for states east of the Great Plains. The COLE program can be accessed at http://ncasi.uml.edu.

The Forest Inventory and Analysis (FIADB) program of the USDA Forest Service provides the most scientifically credible and comprehensive data on the amount and condition of forest resources in the United States. The Forest Inventory and Analysis Database has the potential for a wide array of applications (Miles *et al.* 2001). High-quality tools for online access to the FIADB will assure that this potential is realized. The goal of the Carbon Online Estimation (COLE) project is to develop an online tool to provide access to the FIADB using a versatile user interface (UI) while maintaining a fast response time. With minor revision, the code for COLE will also be useful for providing access to spatial databases and for analyzing the data.

In the first year of development, we have primarily focused on creating an interface and backend that will embody the versatility and speed described above. Additionally, we have created sample queries that calculate carbon levels based on the Eastwide Forest Inventory Database (EWDB) data. The interface has been designed to make it easy to incorporate the FIADB when it becomes available. Due to security concerns, version 1.0 of COLE is linked to a version of the EWDB, which is a forerunner of the FIADB (Hansen *et al.* 1992). We have augmented the EWDB with estimates of carbon for forested plots. It is anticipated that COLE version 2.0 will be linked directly to the FIADB after security issues can be overcome.

Technologies

We want the tool to provide online access with a fast response time while featuring an easy-to-use user interface. The merits of a fast computation time are obvious and easy to quantify, but the idea of a "versatile" UI is less so. UI's are probably the most difficult component of software development. They require the users to familiarize themselves with how each new program works, along with its particular idiosyncrasies. Users can become productive more easily if they are not required to learn a new interface for each piece of software they use. In designing a tool to access the FIADB, it would be fruitful to create a powerful interface that could be easily recycled for various applications. This increases user familiarity and the usefulness of the FIADB.

In approaching the COLE project, the first step was to decide which technologies we would enlist. Our primary goals were to minimize cost and assure operating-system portability. To minimize the cost, the clear choice is open-source software. In today's market, open-source software is a stable and valid option with no licensing cost. We also wanted operating-system portable software that would work on the Windows, Linux, and Macintosh platforms. The solution to this problem was to utilize a cross-platform programming language that could be compiled and/or ported to all three systems.

Keeping these principles in mind, we chose a Linux platform to run the COLE engine. The Linux operating system is a free, stable, open-source platform that is widely recognized as one of the preeminent alternatives to Microsoft Windows and Macintosh. We also needed a database to run from our own

[1]Programmer and Principal Research Scientist, respectively, National Council for Air and Stream Improvement, 600 Suffolk Street Fifth Floor, Lowell, MA 01854. E-mail: paul_vandeusen@uml.edu.

[2]Project Leader, Research Forester, and Research Plant Physiologist, respectively, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, P.O. Box 640, Durham, NH 03824. E-mail: Lheath@fs.fed.us.

local server. We chose MySQL because it is an open-source database query engine compatible with SQL. Since MySQL strictly adheres to the SQL standard, the database we use, and any modifications made to it, can be easily ported to other SQL compatible applications such as Oracle or Sybase.

After establishing our operating and database platforms, we had to decide on a programming language to develop COLE. We wanted a cross-platform language that could be used on all three major operating systems. We also wanted the language to be fast and Web accessible. The two available options were an HTML/CGI/Perl system or a Java platform. Perl offers less than Java in graphical client-side conveniences that would lead to a familiar, easy-to-use interface. Therefore, we selected Java with the accompanying applet/servlet technology. Java is easily compiled on all three major platforms, offers a wide array of UI design tools, and is built to handle Web access. Having decided on a Java platform, we enlisted the open-source Jakarta Tomcat Servlet engine (also available on all platforms) to handle the client-server communications.

User Interaction

We had to select the most logical way for the user to interact with the EWDB. Some existing technologies use drop-down menus and other text-based input to define EWDB queries. However, the EWDB is largely a geographical database, and it would be both intuitive and logical for the users to begin their query by selecting an area on a map.

While the idea of a map-based selection tool was a good one, it was also necessary to decide how such a tool would work. To have a great deal of user power and flexibility, it was necessary to use a map file-format that would allow the users to define their own polygons or use predefined polygons. Furthermore, a display of latitude and longitude would be needed such that the user could easily define accurate and relevant shapes.

Users also want to modify their query with easy-to-use text-based filters and sort variables. Inputs for these factors could be easily designed using the EWDB's predetermined fields. Combining map-based area selection with text-based query modifications provides the user with a powerful query-building tool that is also easy to use.

Development

The first problem in developing COLE was how to make the map interface accessible and dynamic. The obvious answer was to utilize the ESRI Shapefile® format, since the format is publicly available. This allowed us to read shapefiles directly with Java by using the specifications found in the ESRI document. Additionally, the format is widely used and there is a large library of available shape files that could be utilized for COLE.

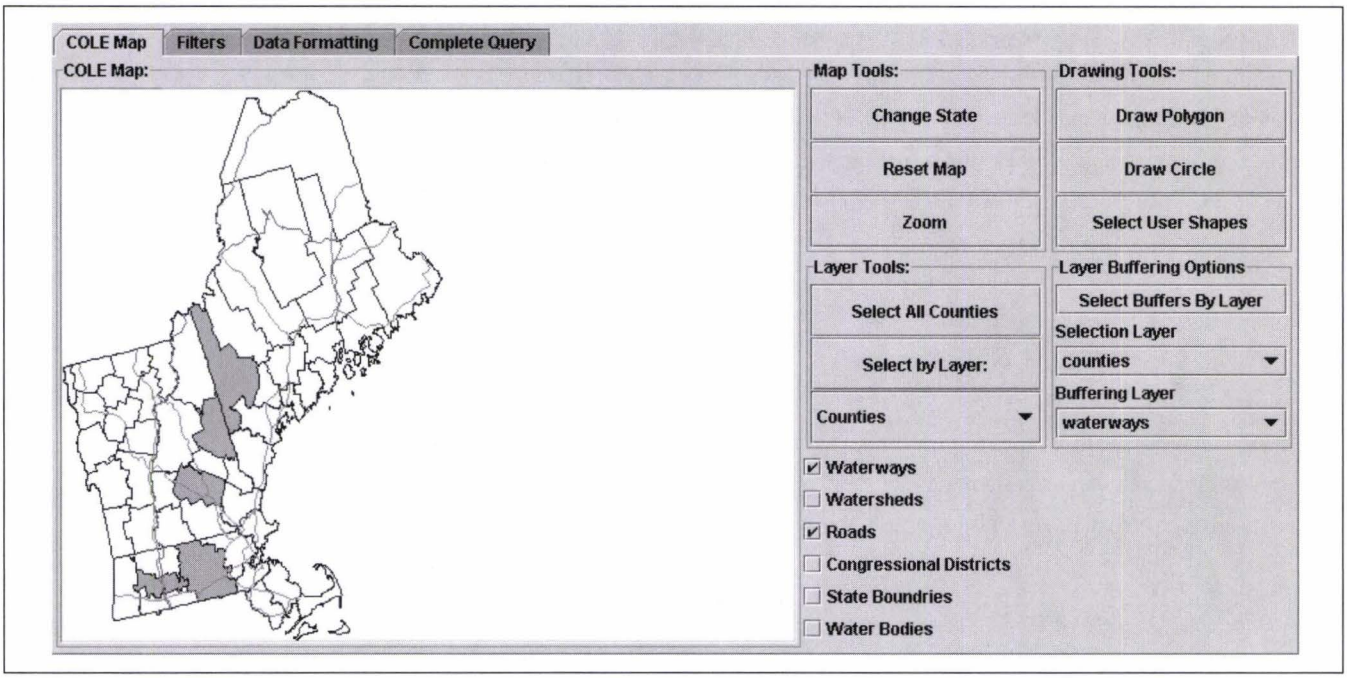
Once we had decided to use the ESRI shape format, we had to make the format interact with Java. While there are available commercial Java/ESRI tools, we decided instead to design our own parser and use it in conjunction with the Java2D Graphics. This would allow us to display the shapes on a Java Canvas object in the interface and enable the user to interact with the map using the properties inherent to Java2D Shape objects. Using these methods, we created a working Canvas-based Java class to facilitate our map display.

With the capacity to display our maps as desired using a standard format, we moved on to defining filters and sorting variables. While the actual sorting and filtering would take place server-side, it was necessary to make a user-friendly representation of those sort and filter variables such that the users could successfully modify their query. To make COLE as flexible as possible, we decided to have the filters and sort variables defined in regular text files. This would allow a COLE administrator to modify the text files and, in turn, modify COLE's capabilities easily. The text files contain both an SQL representation of variables and an English representation. This allows for a link between the user interface and SQL database when submitting a query.

The client side of the program allows the user to pick an area of interest on a map and then submit a query. It is then necessary to develop the server-side mechanism to handle the computations. For our purposes, we created a Java servlet that submits queries to the database, analyzes them for spatial relevance, applies computational algorithms, and returns a data table to the user. By modifying this servlet, one could potentially add more computational functions and increase the number of data formats for the return table.

In completing this server-side functionality, the two primary tasks of the COLE engine are defined. What remains is to build a

Figure 1.— *The tabbed interface of COLE.*



user interface that could leverage these functional technologies and present them to the user in a simple graphical interface.

The COLE Interface

The COLE interface has now reached a reasonably stable and usable layout. The interface enlists the use of Java "tabbed" panes (fig. 1). This allows the user to easily separate submitting a query to COLE into a series of distinct tasks, which combine to generate a dynamic query. COLE currently utilizes a four-tabbed system that addresses the following query-building steps: map-based area selection, data filters, output formatting, and complete query.

The first COLE tab (fig. 1) allows users to select the area relevant to their query. There are several different ways to do this in the interface. First, the user can create a polygon based on coordinate points selected with the user's mouse. Second, the user can create a circle based on a center and a radius. Third, the user can select areas based on predefined polygons imported from a shape file. This selection currently includes counties, watersheds, U.S. Congressional districts, and State boundaries. Finally, the user can select line-based shape files and run a buffering query. The buffering distance is defined on the final tab. Automatically importing a customized shape file from the user is not supported currently; however, users defining their own polygon serves a similar purpose. Following the area selection, the user can then move to the second tab.

The second COLE tab contains the filters defined in a server-side text file. Currently, the interface accommodates up to nine filters. To select a filter, the user must check the checkbox. Once this has been done, the user can select one or many filters from the given filter category. The data will then be filtered in accordance with the user selection.

The third COLE tab allows the user to modify the formatting and data retrieval parts of the query. This includes choosing sort variables (which define table rows and columns), units, query variable, buffering distance, and analysis function to be applied to these data, such as sum and mean. Sort variables, as mentioned above, are based on a server-side text file. Next are the formatting and unit fields. Units currently returned include English and metric. Table formatting and units are both done server-side and require manipulating the servlet to enact further standards. Data computations also are defined server-side. Currently, one can summarize data using sum, average, median, standard error, standard deviation, or sample size (plots or trees). Finally, if users choose to use the buffering feature, they can select the buffering distance on this tab.

On the fourth and final COLE tab, the user can select the format to receive the retrieved data. Currently the options include four types of tables (Jtable, HTML, Spreadsheet, and Tab Delimited) and the graph option. If the users select the graph option, they must then select which type of graphs to retrieve. By checking any combination of the three graph checkboxes, the users can obtain scatter, bar, or pie charts of their data. Once all of these variables have been selected, the user is ready to submit a query. This can be done by selecting the "submit query" button at the base of the fourth tab.

Using the 4-tab interface COLE offers an easy-to-use, progressive interface for accessing a spatial database. The interface has the advantage of being customizable without actually changing compiled code. Furthermore, the types of queries and databases are highly generalized, allowing for later expansion of COLE capabilities.

Carbon Estimates

The carbon estimates in COLE are pre-computed values based on inventory data from the EWDB. Carbon pools are estimated from plot data, such as growing-stock volume and forest type. Aboveground and belowground tree carbon is estimated using the equations presented in Smith *et al.* (2003). The equations for estimating forest floor carbon based on plot data are taken from Smith and Heath (2002). The down dead wood carbon pool is based on an approach similar to that described in Chojnacky and Heath (2002). Forest soil carbon is determined only by forest type and region and is based on estimates in Johnson and Kern (2003). That is, we are using an estimate of soil carbon based on broad regions and forest types of average conditions; previous land use and management were assumed to not affect soil carbon. This approach to estimating carbon is also discussed further in Heath *et al.* (2003).

Data Visualization Tools

The tabular and spreadsheet data retrieval options described above are useful, but COLE provides other options for viewing the data as well. A graphing option allows the user to display results as graphs. A separate, but similar, interface is available to display estimates as maps.

Graphs

By offering graphing options to the user, COLE allows for visual representation of the data. The scatter plot gives details about individual records without revealing what might be considered sensitive information, such as plot location or estimates on land of a specific owner (see fig. 2). The bar and pie options offer a different perspective of the data by giving an easy-to-read relationship between the totals and their sort categories.

When selecting the graphing option, the COLE engine creates graphs based on the user-selected sort variables. The graphs are dynamically built and labeled based on those sort variables, and then returned to the user. This dynamic graph-building ability means that COLE can offer the users a potentially endless selection of sorted graphs based on their area-based query.

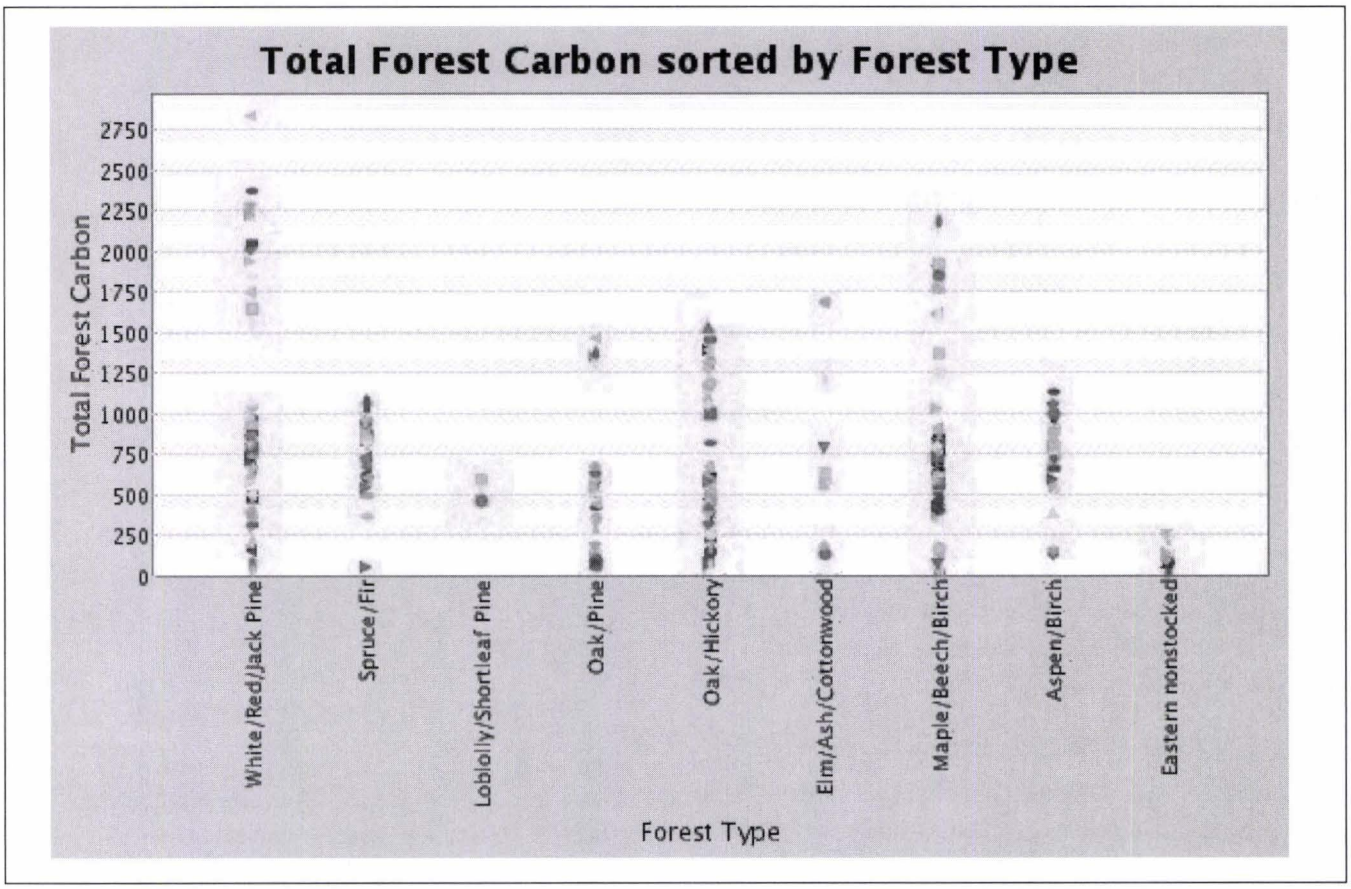
In addition to providing a different perspective on the retrieved data, COLE's graphing option also increases its functionality and usefulness as a data-reporting tool. The graphs are returned in the standard JPEG file format and can be easily saved or cut and pasted in any PC. The resulting files can then be used in presentations, papers, or other media-based interpretations of COLE related data.

COLE-Map

COLE presents a map that defines an area for which data are retrieved, but users also want a tool that would create a map using the retrieved data. This could easily be done with off-the-shelf GIS software but requires more access to the data than most users would be allowed. Adding an online mapmaking capability was the primary goal behind the creation of the COLE-Map as an applet to accompany COLE.

Because COLE was designed as a dynamic toolset, COLE-Map is able to leverage a great many of the technologies used in COLE. COLE-Map utilizes the same tabbed interface as COLE as well as many of the same query-building and submittal classes. The primary differences in COLE-Map occur server-side, and in the absence of any actual map in the user interface.

Figure 2.— *Example scattergraph output from COLE.*



The first difference is largely self-evident. The servlet for COLE-Map performs a different task than that in COLE, and therefore had to be modified. COLE retrieves the data, sorts them, and builds a corresponding table. By contrast COLE-Map retrieves data, sorts them, and then creates and colors a map based on an ESRI shape file. The second difference is also clear: if COLE-Map is making a map, there is no reason for the user to view a map in the interface. Therefore, the map component is removed and replaced with a customizable color palette for the output map. Other than these two minor differences, COLE-Map implements the COLE operations precisely.

Conclusions

In the first year of development, the joint NCASI/USDA Forest Service team has created a powerful and dynamic tool for accessing a version of the EWDB. We have met our initial goal of developing a highly extensible user interface with which to access the EWDB. We have leveraged this user interface to create two separate applications, both of which rely on the extensive data of the EWDB to answer user queries.

This proof-of-concept has opened the door for creating any number of applications of the COLE interface. Other graphical and statistical analysis could be added along with a greater number of GIS-related tools. COLE can readily be linked to additional spatial databases, including the FIADB, with estimation performed on-the-fly. Additional variables, such as the down woody material data from Phase 3 plots, can also be included in linked databases. The usefulness of this tool will only increase as more features are included.

Acknowledgment

This work was made possible by funding from the USDA Forest Service, Forest Health Monitoring program.

Literature Cited

Anonymous. 1998. ESRI Shapefile technical description: an ESRI white paper. Copyright 1997. 34 p.

Chojnacky, David C.; Heath, Linda S. 2002. Estimating down dead wood from FIA forest inventory variables in Maine. Environmental Pollution. 116(1001): S25–S30.

Hansen, Mark H.; Frieswyk, Thomas; Glover, Joseph F.; Kelly, John F. 1992. The eastwide forest inventory data base: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 48 p.

Heath, Linda S.; Smith, James E.; Birdsey, Richard A. 2003. Carbon trends in U.S. forestlands: a context for the role of soils in forest carbon sequestration. In: The potential of U.S. forest soils to sequester carbon and mitigate the greenhouse effect. Boca Raton, FL: CRC Press: 35–45.

Johnson, Mark G.; Kern, Jeffrey S. 2003. Quantifying the organic carbon held in forested soils of the United States and Puerto Rico. In: The potential of U.S. forest soils to sequester carbon and mitigate the greenhouse effect. Boca Raton, FL: CRC Press: 47–72.

Miles, Patrick D.; Brand, Gary J.; Alerich, Carol L.; *et al.* 2001. The forest inventory and analysis database: database description and users manual version 1.0. Gen. Tech. Rep. NC-218. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 130 p.

Smith, James E.; Heath, Linda S. 2002. A model of forest floor carbon mass for United States forest types. Res. Pap. NE-722. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 37 p.

Smith, James E.; Heath, Linda S.; Jenkins, J. C. 2003. Forest volume-to-biomass models and estimates of mass for live and standing dead trees of U.S. forests. Gen. Tech. Rep. NE-298. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 57 p.

The Sensitivity of Derived Estimates to the Measurement Quality Objectives for Independent Variables

Francis A. Roesch[1]

Abstract.— The effect of varying the allowed measurement error for individual tree variables upon county estimates of gross cubic-foot volume was examined. Measurement Quality Objectives (MQOs) for three forest tree variables (biological identity, diameter, and height) used in individual tree gross cubic-foot volume equations were varied from the current USDA Forest Service Forest Inventory and Analysis specifications in a simulation under alternative error models. Assuming unbiased errors may lead to a different control strategy than assuming unbiased errors. Strengthening the MQO for diameter was shown to help reduce the overall variance of volume estimates if diameter errors are slightly biased. Height errors responded favorably to increased control under both the biased and unbiased models. County volume estimates are somewhat robust to the MQOs for biological identity. However, increased control of biological identity did play a more important role when the underlying distributions for diameter and height were assumed to be biased than when these errors were assumed to be unbiased.

The five USDA Forest Service Forest Inventory and Analysis units (FIA) have adopted a common forest inventory design, including core variables, analysis procedures, and quality standards (USDA 2002). An important part of this national effort has been defining Measurement Quality Objectives (MQOs) including acceptable measurement error (or tolerances) for data collected on field plots. Little or no hard data were available to support the initial development of most of these MQOs. So rather than defining the MQOs to achieve a specified maximum variance due to measurement error, they were defined as the best guess as to what might be the specifications achievable by a well-trained observer.

Derived estimates are often the most important factors in considering the utility and applicability of inventory results for a particular purpose. If we wish to control the quality of derived estimates in forest inventories, we must do so by defining measurement quality objectives (MQOs) for those measured variables that contribute information to the derived estimate. To do this, we need to understand the relationships of the error distributions of the measured variables to the error distribution of the derived estimates, and these relationships are typically complex. This paper shows how one may use a simulation to evaluate the contribution to the mean squared error of a derived variable by the allowed error in measured independent variables. In an example, the measurement error allowed by the FIA's existing Measurement Quality Objectives (MQO) for three forest tree variables (species, diameter, and height) used in individual tree gross cubic-foot volume equations were varied in a simulation to examine the effects of the MQOs upon county estimates of gross cubic-foot volume. The simulations were run under two sets of assumptions for two of the variables, height and diameter. I first assumed that the true underlying error distribution was unbiased for each of these variables and subsequently assumed that the true underlying distributions for height and diameter were both biased and skewed.

Assume we are interested in a county attribute mean per acre for county j:

$$Y_j = \frac{1}{L_j} \sum_{i=1}^{N_j} y_i$$

where: Nj equals the number of trees within county j, Lj equals the land area in acres within county j, and yi equals the value of an attribute of tree i. N is uniquely partitioned into N_G groups, $g = 1,\ldots,N_G$. For each group there is a unique function of an easily measured variable vector $\mathbf{x}$ to y:

[1] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, P.O. Box 2680, Asheville, NC 28802–2680. E-mail: froesch@fs.fed.us.

$$y_{i(g)} = f_g(\mathbf{x}_i), \qquad \mathbf{x} = x_{1i}, x_{2i}, x_{3i}, \ldots\ldots$$

Assume further that our variable of interest is gross cubic-foot volume (*gcv)* because it is a pivotal quantity at the FIA unit in the Southern Research Station (SRS), as it enters into equations for most other volume estimates. The *gcv* equations for trees by species group are linear functions of the form:

$v_{i(g)} = a_g + b_g d^2_{i(g)} h_{i(g)}$

where: $v_{i(g)}$ = *gcv* for tree *i* in species group *g*,

$d_{i(g)}$ = diameter of tree *i* in species group *g* at 4.5' above the ground (d.b.h.),

$h_{i(g)}$ = total height of tree *i* in species group *g*, and

a_g and b_g are regression coefficients for species group *g*.

Note that the functional form of volume equations is one aspect of inventory that is not yet standardized nationally. Therefore, the results of this investigation are directly applicable to equations currently used in the southern United States. However, many volume equations used today contain the Schumacher factor: $b_1 d^{b_2} h^{b_3}$ where b_2 is a parameter usually close to 2.0 and b_3 is a parameter usually close to 1.0. Because the Schumacher factor often has an overriding influence in the equation, it is reasonable to expect similar results if we conducted the same studies using the existing equations at other FIA units. For our purposes, we will assume that the functional relationship is known without error. Therefore, if species is correct, as well as diameter and height, the volume is correct.

Methods

Currently, the MQOs require the data collectors to correctly identify the species of all trees 95 percent of the time, and identify the genera of all trees 99 percent of the time. Note that the biological grouping of species into genera does not exactly match the empirical grouping of species referred to above. Depending on how species are grouped, a biological identity error may or may not affect the volume estimate. In addition, it is required that diameter at breast height be measured to within +/- 0.1 inch per 20 inches of diameter 95 percent of the time. Total tree height must be measured to within +/-10 percent of the true height 90 percent of the time (USDA 2001).

The Simulation

A simulation was used to examine the effects of measurement error allowed by the current as well as alternative MQOs upon county estimates of gross cubic-foot volume per acre (*GCV*). Data from the most recent cycle of the FIA survey measured in South Carolina were used which consisted of five consecutively measured panels. Each panel covered the entire State, and all five panels were measured over a period spanning slightly more than 3 years (1998 to 2001). Assuming the data were measured without error, the "true" *GCV* was calculated for each county *j* (GCV_j). For each set of MQOs, biological identity, diameter and height were randomly perturbed within the defined MQOs and error distribution assumptions. Error was randomly applied to the three volume equation variables (biological identity, diameter, and height) for each tree measured in the survey within the defined MQOs and error distribution assumptions. A small quality assurance (QA) data set from the 2000/2001 Forest Health Monitoring (FHM) field season was used to classify the error distributions under the unbiased and biased assumptions for the error distributions. The *gcv* was calculated from these realizations, and the mean gross cubic-foot volume per acre ( ) was calculated for each county in the State. This was ((GĈV)ng the current MQOs, and the alternative MQOs described in table 1. The specifications for each of the three variables were varied while the current specifications of the other two variables were maintained for comparison. The mean difference (MD), mean absolute difference (MAD), and mean squared differences (MSD) from the county results based on the original "true" data were calculated after 1,000 iterations. Specifically, for the error in each estimator of the county mean: $\varepsilon_j = \hat{Y}_j - Y_j = G\hat{C}V_j - GCV_j$, let *C* equal the number of counties in the State, and form three statistics based on 1,000 iterations:

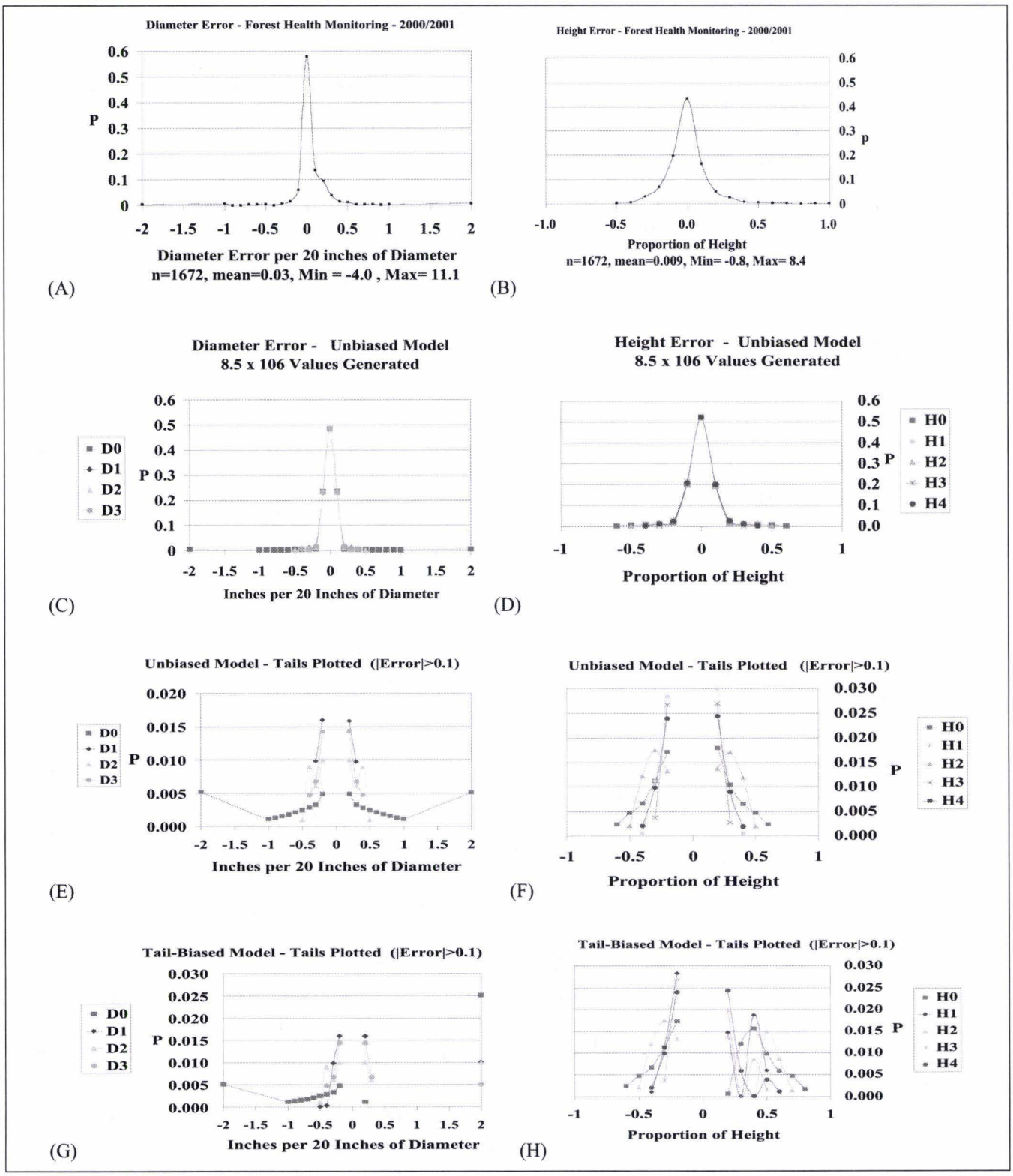
$$MD = \sum_{i=1}^{1000}\left[\frac{1}{C}\sum_{j=1}^{C}\varepsilon_j\right]\Big/1000\,,\; MAD = \sum_{i=1}^{1000}\left[\frac{1}{C}\sum_{j=1}^{C}\left|\varepsilon_j\right|\right]\Big/1000\,,$$

and

$$MSD = \sum_{i=1}^{1000}\left[\frac{1}{C}\sum_{j=1}^{C}\varepsilon_j^2\right]\Big/1000\,.$$

Observations in earlier work (Roesch, in review) and in the 2000/2001 FHM QA data set showed that the error distributions for diameter and height were well behaved in the "in-con-

Figure 1.—*The error distributions observed in the 2000/2001 FHM QA data for diameter (A), and height (B), followed by the simulated distributions for the unbiased model for diameter (C), and height (D). The distribution tails of (C) are rescaled for clarity and plotted in (E). Likewise the tails of plot D are rescaled and plotted in (F). The tails of the simulated distributions for the tail-biased model for diameter are plotted in (G), while the corresponding tails for height are plotted in (H).*



trol" region, and poorly behaved in the "out-of-control" region. So this simulation concentrated on secondary criteria applied to the formally uncontrolled areas of the error distributions.

Species Identity

To vary the species identity determination, a random variate u_i was drawn from a uniform distribution ($U(0,1)$) for each tree i. Let p_1 be the proportion of time that the protocol requires identification of the correct species, and p_2 be the proportion of time that the protocol requires identification of the correct genus, $0 < p_1 < p_2 < 1$. The simulated species determination for tree i, S_i *, was calculated by sampling the following distribution:

$$S_i^* = \begin{cases} S_i & u_i \le p_1 \\ rs(G_{S_i}) & p_1 < u_i \le p_2 \\ rs(F_{S_i}) & p_2 < u_i \le p_2 + r(1-p_2) \\ rs(A) & p_2 + r(1-p_2) < u_i \le 1 \end{cases}$$

where:

S_i = the true species of tree i,

$rs\ (G_{S_i})$ = a random selection from all observed species of the same genus as tree i, except for the species of tree i, unless S_i is the sole species within the genus,

$rs\ (F_{S_i})$ = a random selection from all species in the species list belonging to the same family as tree i, minus those species in the genus of tree i,

$rs\ (A)$ = a random species selection from the entire species list minus those species in the family of tree i, and

r = the proportion of time that out of genus errors are assumed to be within the family of tree i.

Note that under this distribution, the expected value of a correct species call is actually higher than the protocol requires for sole-species genera. This is necessary to meet the within-genus criterion for sole-species genera. The FHM QA data showed that 70 percent of the time when a species identity error fell outside of the correct genus under the current QA specifications, it fell within the correct family. This proportion was used for r in the straw man distribution (S0 in table 1) that is based on the current MQOs. The alternative MQOs for species identity investigated in this study also appear in table 1 (S1, S2, S3, and S4) and involve increases in p_1, p_2 and $p_3 = p_2 + r(1 - p_2)$.

Table 1.—*Alternative additional measurement quality objectives; the units for tolerance (t) are as follows: (1) for species and genus - deviation from true biological identity, (2) for height - the percent deviation from the true height, and (3) for diameter – inches per 20 inches of true diameter. p is the percentage of observations that are required to be correct. The set of alternatives (S0,H0,D0), derived from the original specifications and the 2000/2001 FHM data, form the assumed baseline in table 2.*

Alternative	Specification	*t*	*P*
Current criteria	Species, genus	0, 0	0.95, 0.99
	Height	0.1	0.90
	Diameter	0.1	0.95
S0	Species, genus, family	0, 0, 0	0.95, 0.99, 0.997
S1	Species, genus, family	0, 0, 0	0.975, 0.99, 0.997
S2	Species, genus, family	0, 0, 0	0.975, 0.995, 0.9985
S3	Species, genus, family	0, 0, 0	0.975, 0.995, 0.999
S4	Species, genus, family	0, 0, 0	0.975, 0.995, 0.9995
H0	Height	0.1, 0.5	0.90, 0.99
H1	Height	0.1, 0.2	0.90, 0.95
H2	Height	0.1, 0.3	0.90, 0.95
H3	Height	0.1, 0.2	0.90, 0.98
H4	Height	0.1, 0.3	0.90, 0.99
D0	Diameter	0.1, 1.0	0.95, 0.99
D1	Diameter	0.1, 0.2	0.95, 0.98
D2	Diameter	0.1, 0.3	0.95, 0.98
D3	Diameter	0.1, 0.3	0.95, 0.99

Diameter

Recall that the current specifications require that diameter at breast height is measured to within +/- 0.1 inch per 20 inches of diameter 95 percent of the time. Four alternative specifications (D1, D2, D3, and D4 in table 1) are compared to two straw man distributions based on the current MQOs (D0 in table 1). The first straw man distribution for diameter error is created by splining overlapping unbiased normal distributions, scaled by p_1, t_1, p_2, t_2 and d_{cat}. The unbiased straw man distri-

bution assumes that the small bias observed in the 2000/2001 FHM data is an anomaly of that particular data set and is ignorable. An assumption that diameter measurements are unbiased is supported by the findings of Pollard and Smith (2001). We will use this FHM QA data (plotted in figure 1(A)) to classify the tails of the distribution. The error distributions of this data set might differ from the true underlying distribution because the data are weighted toward inexperienced observers, and they measure observer-to-observer error rather than observer-to-truth error. Let:

z_i = a random variate from a $(N(0,1))$ for tree i,

p_1= proportion of time the measurement must be within a tolerance t_1, of true diameter (d_i), in tenths of an inch per 20 inches of diameter, $0 < p_1 < p_2 < 1$,

p_2 = 0.99, the proportion of time the 2000/2001 FHM data fell within 1.0" (t_2) of the true diameter.

$$d_{cat} = \left[trunc\left(d_i/20.0\right)+1 \right]$$

$$\alpha_1 = 0.5\left(1-p_1\right)$$

$$\alpha_2 = 0.5\left(1-p_2\right)$$

$$e_i = \begin{cases} z_i\left(d_{cat}t_1+0.0499\right)/z_{\alpha_1} & if \quad |z_i| \le z_{\alpha_1} \\ -d_{cat}t_1 - 0.05 + \left(d_{cat}\left(t_2-t_1\right)\frac{\left(z_i+z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ d_{cat}t_1 + 0.05 + \left(d_{cat}\left(t_2-t_1\right)\frac{\left(z_i-z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i > 0\right) \\ -0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ 0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i > 0\right) \end{cases}$$

The distributions we used to compare increased tolerance specifications to the straw man differ from the straw man only in the definitions of p_2 and t_2 and inferences about them. Here t_2 is a required tolerance to be met a proportion p_2 of the time, rather than an assumed parameter of the underlying distribution. That is, we are enforcing a second tier of control, which is more restrictive than the underlying error distribution that arises from the original level of control. Therefore, the error distributions for the alternative distributions are the same as the straw man error distribution, save for the definition and interpretation of p_2 and t_2. Now:

p_2 = the proportion of time the measurement must be within a tolerance t_2, of true diameter (d_i), in tenths of an inch per 20 inches of diameter, $0 < p_1 < p_2 < 1$.

The error distributions for diameter resulting from applying the unbiased error distribution 8.5×10^6 times under the various sets of MQOs for diameter are seen in figure 1(C), while the tails of the distributions are rescaled for clarity in figure 1(E). Note that the long tails observed in the FHM QA data have been retained in the straw man distribution, but they are reduced as control due to the MQOs increases.

An alternative Straw Man distribution would arise if we thought that the previously ignored bias in the 2000/2001 FHM data indicated the true underlying distribution. To model the slight bias and skewness, we would alter our original straw man by applying all of the observed bias to the right tail. Therefore, the second straw man distribution for diameter error is identical to the first except that the bias observed in the 2000/2001 FHM data is added to the right tail of the distribution to approximate both the observed bias and skewness. Let:

$b_i = 0.031 d_{cat}\left(2/(1-p_1)\right)$

Then:

$$e_i = \begin{cases} z_i\left(d_{cat}t_1+0.0499\right)/z_{\alpha_1} & if \quad |z_i| \le z_{\alpha_1} \\ -d_{cat}t_1 - 0.05 + \left(d_{cat}\left(t_2-t_1\right)\frac{\left(z_i+z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ b_i + d_{cat}t_1 + 0.05 + \left(d_{cat}\left(t_2-t_1\right)\frac{\left(z_i-z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right) \cup \left(|z_i| \le z_{\alpha_2}\right) \cup \left(z_i > 0\right) \\ -0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i \le 0\right) \\ b_i + 0.05 + d_{cat}t_2\frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right) \cup \left(z_i > 0\right) \end{cases}$$

For the alternative distributions, we assume that bias can be eliminated from the " in-control" region of the distributions when the second level of control is applied. Therefore, the alternative error distributions are identical to the biased straw man error distribution shown above except that they do not include the bias term in the third line on the right hand side of the equation.

The tails of the error distributions for diameter resulting from sampling the biased error distribution 8.5×10^6 times under the various sets of MQO's for diameter are seen in figure 1(G). That graph shows that the bias, skewness, and influence of the tails are all reduced as the MQOs are increased.

Diameter entries, regardless of error are always interpreted as recorded to the measurement interval of 0.1 inch, and are never negative. Therefore, let *round()* be an operation that

rounds to whole integers. Then all of the diameter error distributions are made discrete to the measurement interval, and negative diameters at set equal to zero:

$$d_i^* = \begin{cases} \left(round\left(10.0*\left(d_i+e_i\right)\right)\right)/10.0 & if\ \left(d_i+e_i\right)\geq 0 \\ 0 & otherwise \end{cases}.$$

The bias to *gcv* estimates added by truncation of diameters at zero is extremely small, since the *gcv* of trees less than 5.0 inches in diameter is zero. Therefore, we will note but otherwise ignore this small amount of bias added to the perturbed diameters.

Height

Because the proportion of height error data from the FHM QA data set showed roughly the same properties as the diameter error data, I used the same approach to simulating height error as I did for diameter error. First, I assumed that the observed bias and skewness are simply anomalies found in that data set rather than indicating the true underlying distribution for FIA proportion of height error. Then, in a second straw man distribution I assumed that the observed bias and skewness truly indicates the underlying distribution. The unbiased model was formed the same way as the unbiased model for diameter. That is by using a spline of overlapping normal distributions, the first scaled by p_1, t_1, and true height and the second scaled by p_2, t_2 and true height. Let:

z_i = a random variate from a $(N(0,1))$ for tree i,

p_1 = the proportion of time the measurement must be within a tolerance t_1, of true height, $0 < p_1 < p_2 < 1$,

$p_2 = 0.99$, the proportion of time the FHM data fell with in 0.5 t_2 of the true height.

$$\alpha_1 = .5\left(1-p_1\right)$$

$$\alpha_2 = .5\left(1-p_2\right)$$

$$e_i = \begin{cases} z_i\left(h_i t_1 + 0.499\right)/z_{\alpha_1} & if \quad |z_i| \leq z_{\alpha_1} \\ -h_i t_1 - 0.5 + \left(h_i\left(t_2-t_1\right)\frac{\left(z_i+z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right)\cup\left(|z_i| \leq z_{\alpha_2}\right)\cup\left(z_i \leq 0\right) \\ h_i t_1 + 0.5 + \left(h_i\left(t_2-t_1\right)\frac{\left(z_i-z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right)\cup\left(|z_i| \leq z_{\alpha_2}\right)\cup\left(z_i > 0\right) \\ -0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right)\cup\left(z_i \leq 0\right) \\ 0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right)\cup\left(z_i > 0\right) \end{cases}$$

As in the case of diameter, the distribution that we used to compare increased tolerance specifications to the straw man differ from the distribution for the straw man only in the definitions of p_2 and t_2 and inferences about them. Here t_2 is a required tolerance to be met a proportion p_2 of the time, rather than an assumed parameter of the underlying distribution. That is, in the alternative MQO specifications (H1, H2, H3, and H4) we are enforcing a second tier of control that is more restrictive than the underlying error distribution that arises from the original level of control (H0). Therefore, the error distributions assuming no bias are the same as the straw man error distribution, save for the definition and interpretation of p_2 and t_2:

p_2 = the proportion of time the measurement must be within a tolerance $t_2 h_i$, of true height (h_i), $0 < p_1 < p_2 < 1$.

The error distributions for proportion of height resulting from sampling the unbiased error distribution 8.5×10^6 times under the various sets of MQOs for height are seen in figure 1(D). The tails of that graph are rescaled for clarity and plotted in figure 1(F). As with the unbiased diameter error distributions, the tails are drawn toward the center as MQOs are increased.

Again, an alternative straw man distribution would arise if we thought that the previously ignored bias and skewness in the 2000/2001 FHM data was somewhat indicative of the true underlying distribution. The second straw man distribution for proportion of height error is identical to the first, except that the small amount of bias observed in the 2000/2001 FHM data is added to the right tail of the distribution to approximate both the bias and skewness seen in that data:

$$b_i = 0.00901 h_i\left(2/\left(1-p_1\right)\right)$$

$$e_i = \begin{cases} z_i\left(h_i t_1 + 0.499\right)/z_{\alpha_1} & if \quad |z_i| \leq z_{\alpha_1} \\ -h_i t_1 - 0.5 + \left(h_i\left(t_2-t_1\right)\frac{\left(z_i+z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right)\cup\left(|z_i| \leq z_{\alpha_2}\right)\cup\left(z_i \leq 0\right) \\ b_i + h_i t_1 + 0.5 + \left(h_i\left(t_2-t_1\right)\frac{\left(z_i-z_{\alpha_1}\right)}{\left(z_{\alpha_2}-z_{\alpha_1}\right)}\right) & if \quad \left(|z_i| > z_{\alpha_1}\right)\cup\left(|z_i| \leq z_{\alpha_2}\right)\cup\left(z_i > 0\right) \\ -0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right)\cup\left(z_i \leq 0\right) \\ b_i + 0.5 + h_i t_2 \frac{z_i}{z_{\alpha_2}} & if \quad \left(|z_i| > z_{\alpha_2}\right)\cup\left(z_i > 0\right) \end{cases}$$

As with diameter, we assume that bias can be eliminated from the "in-control" region of the distributions when the second level of control is applied. Therefore, the distributions arising under the alternative MQO specifications (H1, H2, H3, and H4) are identical to the error distribution above, except that they do not include the bias term in the third line on the right hand side of the equation.

Table 2.—*The mean difference (MD), mean absolute difference (MAD), and mean squared difference (MSD) for each alternative MQO specification in table 1, after 1,000 iterations, under the assumptions of the unbiased and biased straw man error distributions. MD and MAD are in ft³/acre. MSD is in (ft³/acre)²*

Errors in county estimates of gross cubic-foot volume per acre.						
MQO	Unbiased model			Tail-biased model		
alternative	MD	MAD	MSD	MD	MAD	MSD
S0,H0,D0	1.414	8.221	116.404	30.170	30.305	1124.661
S1	0.091	8.063	112.607	28.833	28.997	1043.066
S2	0.444	8.057	112.363	29.204	29.357	1065.213
S3	0.471	8.070	112.606	29.223	29.374	1066.557
S4	0.498	8.063	112.459	29.264	29.414	1068.781
H1	1.374	6.384	69.311	21.958	22.109	610.693
H2	1.406	8.001	109.687	21.965	22.391	659.834
H3	1.328	5.782	56.696	17.014	17.235	381.091
H4	1.394	6.351	68.007	15.427	15.859	336.589
D1	1.237	8.011	110.849	22.229	22.636	676.282
D2	1.222	8.020	110.970	22.270	22.673	678.043
D3	1.242	8.022	111.054	19.881	20.466	571.266

The tails of the error distributions for proportion of height resulting from applying the biased error distribution 8.5×10^6 times under the various sets of MQOs for height are seen in figure 1(H). Again, the bias, skewness, and influence of the tails are all reduced as the MQOs are increased. We assume that height errors are also discrete and not negative. Therefore our simulated heights are calculated as:

$$h_i^* = \begin{cases} round(h_i + e_i) & if\ (h_i + e_i) \geq 0 \\ 0 & otherwise \end{cases}.$$

Results and Conclusions

Table 2 gives the mean difference (MD), the mean absolute difference (MAD), and the mean squared difference (MSD) for each MQO specification in table 1, after 1,000 iterations. The results show that assuming unbiased errors may lead to a different control strategy than assuming bias in the "out-of-control" region. Strengthening the MQO for diameter will help reduce the overall variance of volume estimates if diameter errors are slightly biased in this out-of-control region. Height errors responded favorably to increased control in the current out-of-control region under both the biased and unbiased models.

Volume estimates at the county level are somewhat robust to the MQOs for species identity. However, more accurate species identity did play a more important role when the underlying distributions for diameter and height were assumed to be biased than when they were assumed to be unbiased.

Simulation is useful for investigating the effect of MQOs for independent variables on aggregated dependent variable estimates if reasonable error models can be postulated for the measurement errors of the independent variables. In this case, a small amount of QA data was available that is most likely drawn from a population different from the population of interest. Rather than defining a single distribution for height and diameter errors, intended to represent the underlying population of interest, we defined two for each of these variables that are intended to represent the extremes of the true underlying distributions. Any conclusions that could be drawn from both straw man distributions for a particular variable could be considered robust. However, any conclusion that would only be drawn under one of the straw man distributions should probably be applied more cautiously.

The extension of the methodology used in this paper to other measured and derived variables is straightforward. One simply needs to posit reasonable error models for the measured variables and then simulate attribute variance with those models while observing the effect upon the summary statistics of the derived variables. As applicable quality assurance data becomes more available, they will support a decision to maintain, replace, or refine the error models and MQOs.

Literature Cited

Pollard, J.E.; Smith, W.D. 2001. Forest health monitoring 1999 plot component quality assurance report. Research Triangle Park, NC: U.S. Department of Agriculture, Forest Service, National Forest Health Monitoring Program.

Roesch, F.A. In Review. Testing the sensitivity of county-level volume estimates to the measurement quality objectives for independent variables in individual tree volume equations.

U.S. Department of Agriculture. 2001. SRS field guide version 1.5 (update Jan. 19, 2001). Asheville, NC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Southern Research Station. 311 p.

U.S. Department of Agriculture. 2002. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.6. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis.

Modeling Missing Remeasurement Tree Heights in Forest Inventory Data

Raymond M. Sheffield and Callie J. Schweitzer[1]

Abstract.—Missing tree heights are often problematic in compiling forest inventory remeasurement data. Heights for cut and mortality trees are usually not available; calculations of removal or mortality volumes must utilize either a modeled height at the time of tree death or the height assigned to the tree at a previous remeasurement. Less often, tree heights are not available for trees that were determined to be missed tally trees in an initial inventory. In these cases, a height is available for the current measurement, but the initial tree height must be modeled or estimated. In this paper, we present a procedure for predicting either a time 1 or time 2 height. The procedure uses actual tree height information for a tree collected at time 1 or time 2 if available. Incorporating the relationship between actual tree height to predicted height provides for an adjustment to height equations that do not incorporate site quality and stand parameters.

Missing tree heights are a common occurrence in broad forest inventories. They are relatively rare in an initial installation of inventory plots. Most missing tree heights are encountered during the remeasurement of a previously installed plot where initial inventory (time 1) tree heights are recorded. During the remeasurement of inventory plots, the need to supply a value for a missing tree height arises primarily from the following situations.

1. A previously measured tree is missing at time 2 due to natural mortality or cutting; a height at the time of tree death is needed to compute the tree's volume, which is assigned to mortality or removals.
2. A previously measured tree grows across a merchantability threshold for volume calculation; a height for the tree at the threshold diameter at breast height (d.b.h.) is needed to accurately calculate ingrowth volume.
3. A time 2 inventory tree is determined to have been missed by the previous field crew or the previous height is known to be an error; a time 1 tree height is needed to compute the tree's time 1 volume.

To illustrate the above examples, Forest Inventory and Analysis (FIA) data from a recent inventory in South Carolina were examined. Out of more than 111,000 trees tallied for current inventory estimates, far less than 1 percent required a modeled height. Data recorder software and editing procedures eliminated almost all missing or invalid heights. However, nearly one out of three trees in the remeasurement sample (trees measured 8 years earlier) required a modeled height. Fifteen percent of the trees were cut, 4 percent died from natural causes, 9 percent grew across the growing-stock or sawtimber volume threshold, and 2 percent were survivor or missed trees that required a modeled height at time 1 for various reasons. These examples illustrate that tree heights must be modeled often in operational inventories. The procedures utilized in dealing with missing tree heights must accurately account for a wide range of site and stand parameters. The values used for missing heights in a remeasurement inventory can have a major impact on growth, removal, and mortality volume calculations.

The FIA program is currently developing nationally consistent procedures for collecting and compiling inventory data. While initial compilation efforts have focused on current inventory procedures, the development of remeasurement procedures is underway. Accurate estimates of change (growth, removals, and mortality) will require viable, consistent procedures for dealing with missing tree heights in remeasurement settings. Large changes in tree heights from time 1 to time 2 in remeasurement settings can have severe adverse effects on components of change estimates. In this paper, we present a procedure for dealing with missing tree heights that is simple, allows the use of any tree height model, and produces a modeled tree height in remeasurement settings that is relatively stable relative to the actual measured height of the tree at time 1 or time 2.

[1] Supervisory Research Forester, U.S. Department of Agriculture, Forest Service, Southern Research Station, P.O. Box 2680, Asheville, NC 28802; Research Forester, U.S. Department of Agriculture, Forest Service, Southern Research Station, P.O. Box 1387, Normal, AL 35762.

Methods

The procedures selected for use in FIA inventory computations cannot possibly address all stand, site, and tree parameters because of the diversity encountered across large regions and the Nation. There are many different methodologies utilized by the regional FIA units to model tree heights. We have not assembled a list of those procedures here because no single total height equation or set of equations exists that will perform optimally across all regions. Therefore, the retention of regional FIA methodologies for many computations, including tree height models, is necessary for the immediate future. However, we can implement procedures that improve upon the raw, initial tree heights produced by these equations.

In remeasurement situations, as documented above, we always have some information about each tree that can improve our ability to accurately predict an unknown height. We usually have knowledge of the tree's species, d.b.h. and height, either merchantable length or total height. This information is available for the time 1 inventory or for the time 2 inventory. We can obtain a predicted tree height from a model or equation and use this value unaltered for computations of tree volume and growth. However, unless the model incorporates all site and stand parameters that may impact tree height relationships, abnormal changes in tree height can result. These changes will impact volume and growth computations. One can produce a modeled height that is harmonized with an actual measured height for the tree using the following equation.

$$H_m = \left(\frac{K_a}{K_m}\right) U_m \qquad (1)$$

Where

H_m = final modeled total height
K_a = actual (measured) height
K_m = predicted height for tree with known height
U_m = predicted height for tree with unknown height

This equation may be used to predict a missing tree height at time 2 (cut or mortality tree) or a missing tree height at time 1 (missed tree or erroneous initial tree height). The procedure makes adjustments to raw predicted height values from any equation. The adjustments account for site and stand parameters that have influenced the height development of each individual tree. While stand composition and structural characteristics change due to increasing age and disturbances, the relationship between actual height and modeled height should remain viable for relatively short remeasurement periods of 10 years or less.

Examples

The following examples illustrate the calculation of missing height values for some of the situations described earlier. The examples use data and height equations from the FIA unit at the Southern Research Station (SRS-FIA) to illustrate the effect of the harmonic proportioning equation. D.b.h. and bole length (BL) are the two independent variables in the equations used to derive the initial estimate for a missing height.

$$Y = a + b(d.b.h.) \qquad (2)$$

Trees 1.0 to 4.9 inches d.b.h.

Where

Y = predicted height
d.b.h. = diameter at breast height
a and b are species specific coefficients

$$BL = a + b\left(\sqrt{\log_{10}(d.b.h.)}\right) \qquad (3)$$

Trees 5.0 inches d.b.h. and larger.

Where

BL = predicted length from 1.0-foot stump to 4.0-inch top diameter (outside bark)
d.b.h. = diameter at breast height
a and b are species specific coefficients (table 1)

$$Y = a + b(BL) + c\left(\frac{1}{dbh^2}\right) \qquad (4)$$

Trees 5.0 inches d.b.h. and larger.

Where:

Y = predicted height
BL = bole length
d.b.h. = diameter at breast height
a, b and c species specific total height coefficients (table 2)

Table 1.—*Coefficients for bole length equation by FIA species code*

Species code	a	b	Species code	a	b
131	-156.117206	202.750877	920	-142.109331	185.613178
121	-119.432082	164.99968	621	-148.831031	197.806215
126	-115.074602	152.443899	540	-131.365123	175.827063
128	-124.280736	167.685111	531	-115.243339	151.094033
107	-106.238568	149.364954	370	-85.02753	125.664258
110	-126.330994	169.14767	901	-114.378044	154.378556
111	-152.035297	199.471982	602	-121.652618	161.466442
115	-184.413794	235.884113	491	-26.361267	43.854679
123	-69.543987	109.868908	311	-121.652618	161.466442
132	-106.752125	149.936213	400	-138.431526	179.440462
43	-107.286838	152.696925	591	-89.551605	119.184132
221	-142.478002	186.508392	552	-121.652618	161.466442
10	-130.645957	170.109525	680	-121.652618	161.466442
260	-138.143903	176.104811	521	-121.652618	161.466442
241	-130.645957	170.109525	318	-91.686932	129.421189
222	-134.419057	177.021729	371	-121.652618	161.466442
60	-79.746545	108.851764	837	-121.372298	160.784296
90	-130.645957	170.109525	823	-110.09781	148.608204
129	-135.566314	178.054737	813	-117.535853	162.387287
950	-132.8382	180.305149	832	-94.85546	133.493232
762	-102.309572	139.853199	826	-110.09781	148.608204
694	-150.63215	191.725286	820	-105.69823	143.674925
693	-115.928799	151.979406	838	-107.606289	137.883848
313	-142.109331	185.613178	822	-110.09781	148.608204
330	-142.109331	185.613178	830	-110.09781	148.608204
601	-142.109331	185.613178	835	-95.975318	129.925601
740	-142.109331	185.613178	833	-97.402948	138.559456
651	-96.901447	137.060551	806	-105.862725	146.021142
970	-120.048309	157.758605	817	-110.09781	148.608204
460	-112.391322	149.469203	834	-110.09781	148.608204
555	-108.60678	148.214521	812	-112.741267	149.053737
652	-96.901447	137.060551	825	-116.559834	156.184096
316	-99.705783	137.125344	804	-110.09781	148.608204
580	-142.109331	185.613178	827	-104.639214	143.375171
317	-142.109331	185.613178	802	-109.493036	148.982715
653	-95.979778	136.357591	831	-150.949932	191.23914
611	-161.172361	206.724378	899	-76.443709	103.644195
731	-113.031851	158.006557	999	-87.990247	119.886332
691	-147.46436	193.31287			

The equations use coefficients for 77 species or species-groups in the South. Whereas total stem length is the height value utilized in current volume prediction equations, many of the previous inventories in the South measured merchantable height and modeled total height values from the bole length. Thus, equations are still needed to predict bole length as well as a total height equation utilizing bole length as an independent variable. With total height as the only measured height variable involved in remeasurement plots, a single equation similar to equation (1) can be used to model total heights of trees 5.0 inches and larger.

It should also be noted that d.b.h. is often missing when tree height is missing. For example, both d.b.h. and height are unknown for a tree at the time it was cut. The following examples use modeled d.b.h. values in such instances.

Example 1: Missing Time 1 Tree Height (Missed Tree At Time 1)

Species: chestnut oak (*Quercus prinus*—species code 832)

Time 1 d.b.h.: 19.8 inches (modeled)

Time 2 d.b.h.: 22.0 inches (measured)

Time 2 total height: 72 feet (measured)

Step one: Calculate modeled time 2 bole length (equation 3)

$BL = A + B*SQRT (\log_{10} (d.b.h.))$

$BL = -94.85546 + 133.493232*SQRT (\log_{10}(22.0))$

$BL = 59.813755$

Step two: Calculate modeled time 2 total length (equation 4)

$Y = A + B (BL) + C (1/(d.b.h.)^2)$

$Y = 21.244492 + 0.907202(59.813755) + 141.15111(1/(22.0)^2)$

$Y = 75.799285$

Step three: Calculate modeled time 1 bole length (equation 3)

$BL = A + B*SQRT (\log_{10} (d.b.h.))$

$BL = -94.85546 + 133.493232*SQRT (\log_{10} (19.8))$

$BL = 57.154893$

Step four: Calculate modeled time 1 total length (equation 4)

$Y = A + B (BL) + C (l/(d.b.h.)^2)$

$Y = 21.244492 + 0.907202(57.154893) + 141.15111(1/(19.8)^2)$

$Y = 73.455568$

Step five: Calculate harmonically proportioned total length for time 1 (equation 1)

$H_m = (K_a/K_m)*U_m$

$H_m = (72.0/75.799285) * 73.455568$

$H_m = 69.8$ (final estimate for initial tree height)

In this example, the first estimate of time 1 total height was 73 feet, which was greater than the measured value for the tree at time 2. The proportioning procedure results in a more realistic value that is consistent with d.b.h. change.

Example 2: Missing Time 2 Tree Height (Cut Tree)

Species: longleaf pine (*Pinus palustris*—species code 121)

Time 1 d.b.h.: 6.0 inches (measured)

Time 2 d.b.h.: 8.0 inches (modeled)

Time 1 total height: 54 feet (measured)

Step one: Calculate modeled time 1 bole length (equation 3)

$BL = A + B * SQRT (\log_{10} (d.b.h.))$

$BL = -119.432082 + 164.99968*SQRT (\log_{10} (6.0))$

$BL = 26.118891$

Step two: Calculate modeled time 1 total length (equation 4)

$Y = A + B (BL) + C (1/(d.b.h.)^2)$

$Y = 7.729578 + 1.038906 (26.118891) + 460.44451(1/(6.0)^2)$

$Y = 47.65478$

Step three: Calculate modeled time 2 bole length (equation 3)

$BL = A + B*SQRT (\log_{10} (d.b.h.))$

$BL = -119.432082 + 164.99968*SQRT (\log_{10} (8.0))$

$BL = 37.368841$

Step four: Calculate modeled time 2 total length (equation 4)

$Y = A + B (BL) + C (l/(d.b.h.)^2)$

$Y = 7.729578 + 1.038906 (37.368841) + 460.44451(1/(8.0)^2)$

$Y = 53.74674$

Table 2.—*Coefficients for total tree height equation by FIA species code*

Species code	a	b	c	Species code	a	b	c
131	11.63601	1.002513	347.272658	920	18.718746	0.973277	237.046549
121	7.729578	1.038906	460.44451	621	20.671028	0.956696	219.505156
126	12.235965	0.949229	110.433871	540	20.026148	0.955226	257.081973
128	9.785433	0.99598	265.638401	531	33.050946	0.781059	-203.793929
107	12.138889	0.976357	255.670995	370	26.749454	0.875348	53.60581
110	8.227609	1.036995	345.204061	901	24.201253	0.890276	45.69745
111	7.627238	1.046602	429.084288	602	21.840821	0.899214	166.601854
115	11.612768	0.998582	340.837362	491	15.84294	0.996554	81.45163
123	3.433555	1.062757	308.1748	311	20.91845	0.967155	143.329004
132	11.719363	0.993829	289.806791	400	20.279042	0.978179	232.507527
43	11.865871	1.000782	235.023676	591	15.702126	0.930761	35.053649
221	20.131836	0.9007	201.316112	552	20.91845	0.967155	143.329004
10	14.527042	0.978106	162.61837	680	22.055232	0.86043	-19.111068
260	20.872663	0.905741	7.644922	521	9.245754	1.051793	407.441276
241	14.527042	0.97816	162.61837	318	16.476717	1.03106	307.64589
222	7.987471	1.02906	497.921189	371	24.394062	0.945372	91.247826
60	14.032483	1.003766	122.450047	837	21.297146	0.952646	226.151486
90	14.527042	0.978106	162.61837	823	17.367545	1.006456	197.818822
129	15.330778	0.95947	219.242547	813	16.436721	1.040216	209.859337
950	24.914346	0.89236	38.929822	832	21.244492	0.907202	141.15111
762	15.738263	0.996829	308.494651	826	21.589084	0.933223	-123.979563
694	15.403016	0.998625	246.855337	820	16.78362	0.985803	238.462787
693	16.746882	0.963007	146.246985	838	17.42442	0.903906	116.917948
313	18.718746	0.973277	237.046549	822	17.367545	1.006456	197.818822
330	9.731165	1.146179	401.643379	830	17.367545	1.006456	197.818822
601	18,718746	0.973277	237.046549	835	20.755906	0.924214	64.282713
740	18.718746	0.973277	237.046549	833	17.354367	1.023787	244.745485
651	16.866377	1.009975	299.861067	806	20.636356	0.964852	128.255717
970	26.041732	0.872665	-6.198928	817	17.367545	1.006456	197.818822
460	31.13191	0.810872	-188.932025	834	17.367545	1.006456	197.818822
555	14.664977	0.990312	242.687846	812	17.40684	0.993855	236.337343
652	13.031815	0.990918	475.180647	825	24.489813	0.891002	10.363523
316	23.972077	0.882021	129.85142	804	17.367545	1.006456	197.818822
580	18.718746	0.973277	237.046549	827	20.649189	0.947504	139.682441
317	18.718746	0.973277	237.046549	802	20.445897	0.961278	148.494802
653	17.455189	0.927531	240.838341	831	19.57347	0.992177	207.241184
611	20.336715	0.950014	235.637303	899	12.147201	1.015202	122.112775
731	25.513236	0.894188	107.211601	999	18.951285	0.920112	122.341224
691	19.550482	0.951712	363.668158				

Step five: Calculate harmonically proportioned total length for time 1 (equation 1)

$H_m = (K_a/K_m)*U_m$

$H_m = (54.0 / 47.65478) * 53.74674$

$H_m = 60.9$ (final estimate for terminal tree height)

This example illustrates the value of using the harmonic proportioning procedure to adjust initial predicted values for site and stand conditions. The first estimate of total height at time of cutting was 54 feet, an average value for the species and d.b.h. If we used this value, there would be no height growth recorded for the tree. However, the relationship between actual height and predicted height at time 1 provides a site- and stand-specific ratio to use in producing a more likely height value of 61 feet.

Example: Calculate Probable Height at Time of Ingrowth

Species: yellow-poplar (*Liriodendron tulipifera*—species code 621)

Time 1 d.b.h.: 2.5 inches (measured)

Time 2 d.b.h.: 6.8 inches (measured)

Time 1 total height: 24 feet (measured)

Time 2 total height: 49 feet (measured)

In this example, we have a known d.b.h. and total height at both time 1 and time 2; we want to obtain a height when the tree crossed the merchantability threshold for volume calculation, which is 5.0 inches d.b.h. We could use the time 1 or time 2 actual values in the calculation of the proportion—we use time 2 values here.

Step one: Calculate modeled time 2 bole length (equation 3)

$BL = A + B * SQRT (\log_{10} (d.b.h.))$

$BL = -148.831031 + 197.806215*SQRT (\log_{10} (6.8))$

$BL = 31.651170$

Step two: Calculate modeled time 2 total length (equation 4)

$Y = A + B (BL) + C (1/(d.b.h.)^2)$

$Y = 20.671028 + 0.956696 (31.651170) + 219.505156 (1/(6.8)^2)$

$Y = 55.69866$

Step three: Calculate modeled bole length at ingrowth d.b.h. (equation 3)

$BL = A + B*SQRT (\log_{10} (d.b.h.))$

$BL = -148.831031 + 197.806215*SQRT (\log_{10} (5.0))$

$BL = 16.543720$

Step four: Calculate modeled total length at ingrowth d.b.h. (equation 4)

$Y = A + B (BL) + C (1/(d.b.h.)^2)$

$Y = 20.671028 + 0.956696 (16.543720) + 219.505156 (1/(5.0)^2)$

$Y = 45.27854$

Step five: Calculate harmonically proportioned total length at ingrowth d.b.h. (equation 1)

$H_m = (K_a/K_m)*U_m$

$H_m = (49.0 / 55.69866) * 45.27854$

$H_m = 39.8$ (final estimate for tree height at ingrowth d.b.h.)

The tree in this example has achieved height growth well less than average based upon the relationship of predicted versus actual height at time 2. As a result, the modeled height values at the time of ingrowth into the merchantable volume category are adjusted downward accordingly.

Discussion

The harmonic proportioning procedure presented above has been operational in remeasurement inventory processing procedures at SRS-FIA for several years. It is effective in preventing unnecessary fluctuations in modeled tree heights and resulting volumes. The procedure is easily incorporated into existing height calculations and works with any equation or model that provides an initial estimate for a missing height. It effectively accounts for much of the influence that site and stand conditions have on tree height if remeasurement cycles are not excessively long.

An Alternative View of Some FIA Sample Design and Analysis Issues

Paul C. Van Deusen[1]

Abstract.—Sample design and analysis decisions are the result of compromises and inputs from many sources. The end result would likely change if different individuals or groups were involved in the planning process. Discussed here are some alternatives to the procedures that are currently being used for the annual inventory. The purpose is to indicate that alternatives exist and that reasonable people might prefer approaches that differ from the ones selected for the annual inventory. The topics covered include panel creep, mapping, the moving average, and data security issues.

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service is implementing an annual forest inventory (USDA Forest Service 1999) where a percentage of the plots are measured each year. Before the 1998 Farm Bill that mandated the annual inventory, FIA was conducting periodic surveys in each State. Each survey took 1–4 years, and then the State was revisited every 10–18 years. The 1998 Farm Bill followed two Blue Ribbon Panel reports, BRP I and BRP II (American Forest Council 1992, American Forest and Paper Association 1998). BRP I called for shortening the period between surveys from 10 to 5 years. This goal was never achieved, and cycles averaged 10 years or more when BRP II convened in 1997. The BRP II request for more timely data motivated the 1998 Farm Bill legislative mandate for an annual survey. The Farm Bill mandate implies that it is important to have annual updates and continuous data collection underway in all States. It remains to be seen whether this goal will be achieved.

Panel Creep

The interpenetrating panel design was selected because it seemed to best facilitate the intent of the Farm Bill, which was to have a regular proportion of the plots measured each year. The Farm Bill also intended that annual updates should be possible. Presumably, the purpose of annual data collection is to provide annual estimates. Hence, the original plan was to measure one panel each year. A panel consists of plots that systematically cover the State. This means that unbiased estimates of any quantity are possible from a single panel or from combinations of panels. There are also secondary benefits of the panel design, such as maintaining equal annual budgets and work loads.

The initial plan was to create 5 panels in Eastern States and 10 panels in the West. Unfortunately, resources are not available to measure one panel per year in all States. This leads to panel creep. Panel creep occurs when a panel has been only partially measured at the end of the field season. One could differentiate between planned and unplanned panel creep, but the end result is the same. A panel can't be used for a statewide update until all plots have been measured. One way to avoid panel creep is to create subpanels, which amounts to dividing a main panel into smaller panels that each systematically covers the State. The 10 panels out west could be viewed as a 5-panel system with subpaneling. A 5-panel system has 20 percent of the plots in each panel, while a 10-panel system has 10 percent of the plots per panel. Therefore, panel creep in a 10-panel system results in fewer plots being unavailable than in a 5-panel system.

A 20-subpanel design seems like a reasonable compromise. Each panel contains 5 percent of the plots, so leaving a panel uncompleted at the end of the field season is less damaging. At the beginning of the season, the crews could begin measuring as many panels as they expect to complete, say four. It might become clear that only three panels can be finished as the season progresses. A solution is to stop work on one panel so that three full panels are completed. A 20-subpanel design works well when the real goal is to measure either 5 or 10 percent of the plots each year. The current plan to complete a State in 7 years in the East doesn't fit a 20-subpanel design perfectly, but it is close. Measuring 3 subpanels out of 20 each year would be a good approximation.

[1]National Council for Air and Stream Improvement, 600 Suffolk Street, Fifth Floor, Lowell, MA 01854. E-mail: Pvandeusen@ncasi.org.

Mapping Versus Fuzzing

FIA sample plots will not always fall into a uniform forest condition and may straddle multiple conditions. At one time, FIA rotated "straddler" subplots into a uniform forest condition. This created a bias against forest conditions that were long and narrow, and procedures to replace plot rotation were developed (Birdsey *et al.* 1995). The procedure that FIA has officially adopted is referred to as "mapping." This involves mapping the boundary of each distinct condition class that occurs on a subplot, subject to certain limitations. An alternative approach, called fuzzing, is to assign each subplot to a single condition class.

An FIA plot consists of four 1/24-acre subplots in a fixed configuration, and there is one plot for every 6,000 acres. The official justification seems to be that mapping subplots results in the least amount of bias and variance for estimates of condition class volume and area, and this may be true when condition boundaries are well defined. The fuzz method involves classifying each subplot into a single condition class. There are advantages and disadvantages of fuzzing versus mapping (Hahn *et al.* 1995). A major advantage is that field procedures are simple for fuzzing, because no decisions about type boundaries are required.

With fuzzing, a subplot is classified into whatever category occupies the largest proportion of the subplot. Usually, there are no more than two classes per subplot. Suppose we are dealing with pine and hardwood classes and there are several subplots that are 50 percent in each class. Presumably, the field crews would on average call half of these subplots pine and the other half hardwood, which means the end result is unbiased. In any event, a subplot is only 1/24 of an acre, and it seems that splitting it into smaller areas via mapping is unnecessary. The tradeoff in precision may be worth the savings in field-work time, and complexity. This could be formally evaluated with existing mapped data, because the corresponding fuzzed results can be determined from the more detailed mapped data. Such an evaluation would require assuming that the mapped boundaries are distinct and correctly mapped.

The Moving Average

The data from an interpenetrating panel design can be analyzed in many ways, which is a strength of the design. FIA has chosen the 5-year moving average (MA) as the default procedure. A comparison among the MA, a one-panel mean and a mixed estimator (Van Deusen 2002) indicates that the MA works well when there is no trend in the data, but can show significant bias in the presence of a trend.

A brief review of the 5-year moving average will help clarify the problem with bias. The MA, as envisioned for use by FIA, is equivalent to averaging all plot measurements from the last 5 years in a State. For years t-4 through t, this can be written as:

$$MA_{t-4,t} = \sum_{j=t-4}^{t} w_j \bar{y}_j$$

where [$\bar{y}_j$][2] is the average of all plot values measured in year j, and w_j is a weight such that [$\sum w_j = 1$][3] . The weight, w_j, ensures that each panel is weighted according to the proportion of the total plots it contains. With an exact 20 percent sample, [w_j=0.2.] When [$\bar{y}_j$][4] represents a single panel mean for year t, it is unbiased for the true underlying value, [β_t][5] , and we can write [$\bar{y}_t = \beta_t + e_t$][6]

where e_t is a random error term. It follows that the expected value of the moving average is

$$E(MA_{t-4,t}) = \sum_{j=t-4}^{t} w_j \beta_j$$

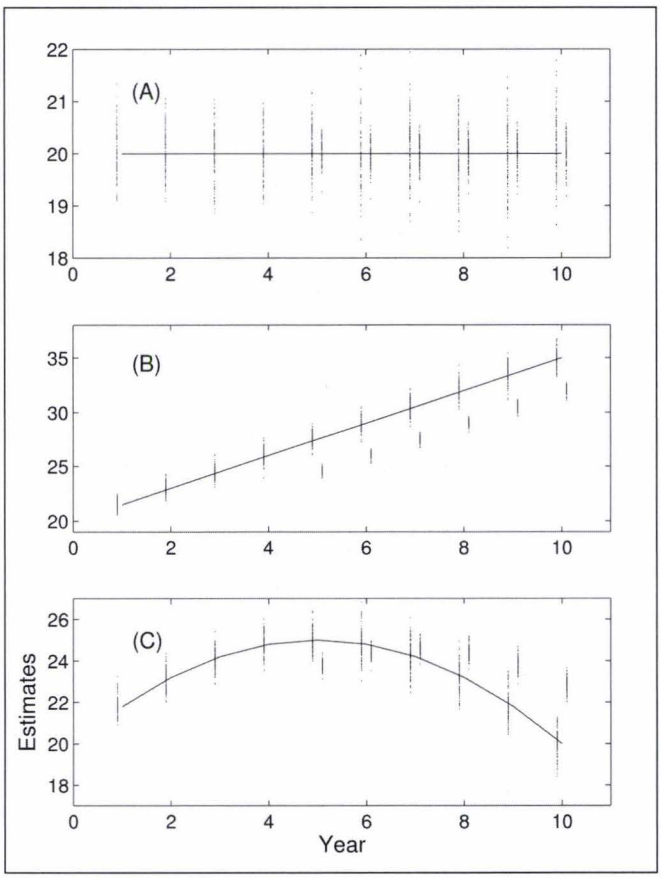
Therefore, [$MA_{t-4,t}$] estimates the true average over the last 5 years and is a biased estimate of the current value, [β_t][8]. If a user wants to estimate the midpoint of a 5-year period, the bias of the MA would generally be less. Most users likely want to be able to associate estimates with particular years, which is the reason an annual inventory received so much support from users.

A comparison using simulated data described elsewhere (Van Deusen 2002) shows the one-panel mean (M1) and the MA (fig. 1) for flat, linear and quadratic trends. The figure displays results from 100 simulations for years 1 through 10. The MA results can be shown only for years 5 through 10, since

they require 5 years of data. The MA is clearly superior when there is no trend. However, it's not clear that the MA is superior for the linear trend since the bias is so large. The quadratic trend shows that the sign of the bias of the MA changes after the trend changes direction.

It is clear that the moving average can show significant bias. Other estimators that can follow trends and yield less bias could be implemented by FIA. The MA also has some weaknesses for estimating change. The bias in a linear trend would be subtracted out when estimating change, but the bias in a quadratic trend would be amplified. Research exists on alternative model-based methods and imputation procedures (McRoberts 2001, Reams and Van Deusen 1999, Roesch and Reams 1999, Rubin 1987), but this is an area where more research may be justified.

Figure 1.—*Comparison of the one-panel mean and the 5-year moving average with three trends. Estimates made with simulated data and 100 repetitions are shown for years 1-10 for the one-panel mean and years 5-10 for the moving average. The trends are (A) flat, (B) linear, (C) quadratic.*



Data Security Issues

There is an inherent conflict between making FIA data available to users and maintaining data confidentiality. Data security places limits on the use of FIA data that can diminish its value, but also prevents potential abuses. FIA must consider private property owners' concerns about these data. To obtain the data, FIA field crews must have permission from landowners to access private land where most FIA plots are located. These private owners are not obligated to grant access, and in most cases derive little direct benefit from FIA activities. Owners are concerned about how the information being collected on their land is being used and who has access to it.

The long-term viability of the FIA program depends on striking a balance between adequate security and providing access to users. FIA access refusal rates are somewhat less than 2 percent nationally. If owners lose confidence in FIA data security, refusal rates will go up. At worst, the FIA program could become ineffective. Likewise, FIA will lose political support if users have too little access to the data.

Data security issues were less important before data became easily accessible in digital form. This situation has changed dramatically over the past two decades with the rise of personal computers and the Internet. The first official indication of this trend may have been in section 1770 of the 1985 Food Security Act (FSA), which creates a legal mandate for USDA employees to protect the identity of individuals who provide data. The 1985 FSA had no immediate impact on FIA data confidentiality procedures, but in 1989 FIA made it a policy to release only fuzzed coordinates to the general public. The fuzzed coordinates were adjusted to be within ±1 mile of the true location. The 1985 FSA became applicable to FIA data following a 1999 amendment (H.R. 3423) that inserted a new paragraph including the Forest and Rangeland Renewable Resources Act.

The 1989 system of fuzzed coordinates was deemed to be inadequate in 2001 due to H.R. 3423, and FIA stopped releasing coordinates to the public. Before this, there were publicly available Eastwide and Westwide databases containing fuzzed coordinates. In the future, data released to the public will be aggregated, fuzzed, or rearranged so that it will not be possible to determine who owns the land that contains a particular plot

or group of plots. These new restrictions may hinder some traditional industry uses of the data, such as mill supply or location surveys. They will certainly hinder research uses of FIA data that require owner and land use information. FIA needs to develop a policy that will protect private owners' rights while not alienating traditional users by unnecessarily limiting access to the data. It remains to be seen if FIA can accomplish such a difficult compromise.

Summary

FIA is still in the process of replacing the periodic inventory with an annual inventory, and many decisions must be made regarding implementation details. It is often true that an organizational decision involves considering competing options with no clearly superior choice. Regardless, the attributes of each option are weighed and a decision is made. It is clear that another group or decisionmaker might have reached a different conclusion. This paper is written in the spirit of acknowledging that reasonable people can reach different conclusions. Likewise, some of the initial decisions made when implementing the annual inventory could change as new information and viewpoints become available. The issues discussed here were panel creep, mapping versus fuzzing, the moving average, and FIA data security.

Panel creep will almost certainly occur within the annual inventory system. The discussion of panel creep should be about minimizing the impact of not being able to use plot data until a panel is completed. FIA chose to implement 5 panels in the East and 10 in the West, even though few Eastern States would be able to measure 1 panel per year. This ensures that up to 20 percent of already measured plots may be unavailable in the current year. One solution would be to create additional subpanels to decrease the number of unusable, measured plots.

Mapping is a fix for the bias caused by plot rotation. The issue here is that mapping might be unnecessarily detailed and time consuming and that a method known as fuzzing could be sufficient. Fuzzing might result in more bias than mapping, but it would simplify field work and analysis.

The moving average was chosen as the default estimator because it is easy to use. The issue here is that ease of use comes at the expense of bias. In this case, we see FIA accepting bias as a tradeoff for ease of use, whereas bias associated with fuzzing was not accepted. This seems to demonstrate the contention that different groups can reach different conclusions.

The final decisions related to the issue of data security had not been made as this paper was being written. Whatever they turn out to be, there is little doubt that reasonable people will disagree about them. Some will feel they went too far and prevented legitimate uses of the data and others will feel they were too lenient.

Literature Cited

American Forest and Paper Association. 1998. The report of the second blue ribbon panel on forest inventory and analysis. Washington, DC.

American Forest Council. 1992. The report of the first blue ribbon panel on forest inventory and analysis. Washington, DC: American Forest Council.

Birdsey, R.A.; Hahn, J.T.; MacLean, C.D.; *et al.* 1995. Techniques for forest surveys when cluster plots straddle two or more conditions. Forest Science Monograph 31. 82 p. Supplement to Forest Science 41(3).

Hahn, J.T.; MacLean, C.D.; Arner, S.L.; Bechtold, W.A. 1995. Procedures to handle inventory cluster plots that straddle two or more conditions. Forest Science Monograph 31. 82 p. Supplement to Forest Science 41(3).

McRoberts, R.E. 2001. Imputation and model-based updating techniques for annual forest inventories. Forest Science. 47: 322–330.

Reams, G.A.; Van Deusen, P.C. 1999. The southern annual forest inventory system. Journal of Agricultural, Biological, and Environmental Statistics. 4(4): 346–360.

Roesch, F.A.; Reams, G.A. 1999. Analytical alternatives for an annual inventory system. Journal of Forestry. 12: 33–37.

Rubin, D. B. 1987. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley and Sons. 258 p.

U.S. Department of Agriculture. 1999. A strategic plan for forest inventory and monitoring. Washington, DC: Department of Agriculture, Forest Service. 48 p.

Van Deusen, P.C. 2002. Alternative sampling designs and estimators for annual surveys. In: Hansen, M.; Burk, T., eds. Proceedings, Integrated tools for natural researches inventories in the 21st Century, proceedings of the IUFRO conference; 1998 August 16–20; Boise, ID. Gen. Tech. Rep, NC–212. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station: 192–196.

Van Deusen, P.C. 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. Canadian Journal of Forest Research. 27: 379–384.

Sensitivity of FIA Volume Estimates to Changes in Stratum Weights and Number of Strata

James A. Westfall and Michael Hoppus[1]

Abstract.—In the Northeast region, the USDA Forest Service Forest Inventory and Analysis (FIA) program utilizes stratified sampling techniques to improve the precision of population estimates. Recently, interpretation of aerial photographs was replaced with classified remotely sensed imagery to determine stratum weights and plot stratum assignments. However, stratum weights based on remotely sensed data depend on many factors, such as classification algorithm and image selection. County volume estimates and associated variances were calculated over a range of stratum weight scenarios and for various number of strata. Rates of change in estimated values and variances, and their effects on percent sampling error, were examined in relation to different strata configurations.

Historically, the USDA Forest Service Forest Inventory and Analysis (FIA) program has used stratified estimation techniques to increase the precision of estimates of forest population parameters. Stratifying the population has traditionally been done by interpreting aerial photography. However, advances in technology now favor the use of satellite imagery for many of the regional FIA units. In the Northeast (NE) region, stratification is performed with data from the Landsat 7 sensor platform. Each pixel in an image is classified and the resulting classes are collapsed into groups that represent different strata. Information on sample-plot location is used to assign plots to strata, and stratum weights are determined by stratum size in relation to total area. The computed stratum weights depend on several factors. Images of the same land area taken at different times produce different stratification results. Also, many algorithms can be used to classify an image. Each algorithm produces different classification results, which affect the stratification outcome. As stratification results are utilized in the estimation process, it is of interest to investigate the degree to which different results might affect estimated values. In our study, we examined the extent to which changes in stratum weights and number of strata affect estimates of merchantable (4-inch top limit) cubic-foot volume and associated sampling errors.

Data

Data for this study are from a new FIA annual inventory system (USDA Forest Service 2002) in Maine, where data exist for about 60 percent of the sample plots. Each plot consists of four circular subplots, each with a radius of 24 feet (7.3 m). Sample plots were mapped in detail by land condition, allowing for an estimate of the area for each condition. Different conditions were delineated among forest, nonforest, water, and other variables, and boundaries were established within forested conditions for other types of changes, such as forest type, stand size, and tree density. Information collected included tree location, diameter at breast height, total and merchantable heights, percent of cull, and type and location of tree damage. Gross cubic-foot volumes to a 4-inch (10.2 cm) top limit for individual trees were computed with equations developed by Scott (1981). The volume of cull is subtracted from the predicted gross volume to obtain net volume. Net plot volumes were obtained by summing volumes of individual trees.

Methods

To estimate total cubic-foot volume for a county, weights for each stratum must be determined. This was done by reclassifying a Landsat TM-based forest/nonforest cover map (Vogelmann *et al.* 1998) using a 5 x 5 moving-window summarization. This placed each pixel into one of 26 classes (0 to 25 forested pixels in the surrounding 5 x 5 pixel box). These class-

[1] Research Forester and Supervisory Forester, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA 19073. Phone: 610–557–4043; fax: 610–557–4250; e-mail: jameswestfall@fs.fed.us.

es were then collapsed into strata and stratum weights by county were computed from map pixel counts and census-derived county boundary information (U.S. Census Bureau 2001).

Using the stratum weights derived from the satellite-image analysis, we calculated net cubic-foot volume estimates and variances by county. The estimation procedures were based on methodology described by Cochran (1977) for stratified random sampling. These estimators utilize the stratum weights to reflect the relative influence of each stratum. Having obtained the estimate and associated variance of the estimate, one can compute the percentage sampling error for the estimate of total volume:

$$SE\% = \frac{\sqrt{v\left(\hat{Y}\right)}}{\hat{Y}}(100) \quad (1)$$

where:

$v\left(\hat{Y}\right)$ = variance of the estimate of total cubic-foot volume
$\hat{Y}$ = estimate of total net cubic-foot volume

To assess how different stratum weights may affect county estimates, estimates and sampling errors for total cubic-foot volume were computed over a range of stratum weights. For each stratum, weights were systematically altered to ±0.25 of the original value in increments of 0.01. Weights of the remaining strata were increased or reduced proportional to their size to maintain the requirement that the sum of all stratum weights equals 1. For example, if the original weight of stratum 1 was 0.30, then the range of stratum 1 weighting scenarios covered 0.05 through 0.55 in increments of 0.01. The ±0.25 range was not fully realized for strata with original weight values less than 0.25 or greater than 0.75, as this would produce weights less than 0 or greater than 1.

Different numbers of strata were created by altering the four original strata. When the number of strata was to be reduced, strata were combined. For this analysis, strata 3 and 4 had similar attributes and were combined to reduce the number of strata to three. To increase the number of strata, an existing stratum was divided into two parts. Because stratum 4 was nearly twice the size of the other strata, this stratum was divided into two equal parts. Half of the plots from the original stratum 4 having the smallest plot volumes were retained as stratum 4, and stratum 5 was created with the remaining plots. Interestingly, this division created five strata with nearly equal weights.

Analysis

The analysis was conducted for Androscoggin County in Maine, a small county (~318,000 ac) that is about 70 percent forested. The cubic-foot volumes and stratum weights used in this analysis are given in table 1. For the original configuration of four strata, the effects of changing weights of various strata on the estimate of volume are depicted in figure 1. The estimates are most sensitive to weight changes for stratum 1, where a 0.01 increase in weight decreased the estimated value roughly 1 percent. Stratum 2 is largely insensitive to changes in stratum weight. Strata 3 and 4 behave similarly, where increases in the estimated volume are nearly 0.5 percent for every 0.01 unit of change in weight. Similar trends are noted for changes in the variance of the estimate (fig. 2), although the magnitude of change is less. The exception is stratum 2, where, unlike the negligible change in the estimate, a notable trend in variance is

Table 1.—*Summary of stratum attributes for Androscoggin County, Maine*

3 Strata	Mean plot volume (ft^3)	Variance of mean plot volume	Stratum weight
1	24.1	442.4	0.226
2	149.2	1,060.3	.180
3	217.2	1,096.0	.594
4 Strata			
1	24.1	442.4	.226
2	149.2	1,060.3	.180
3	233.0	2,927.1	.228
4	204.5	1,864.9	.366
5 Strata			
1	24.1	442.4	.226
2	149.2	1,060.3	.180
3	233.0	2,927.1	.228
4	87.0	719.5	.183
5	321.9	776.3	.183

Figure 1.—*Percent change in estimate of total volume when stratum weights are changed for four strata in Androscoggin County, Maine.*

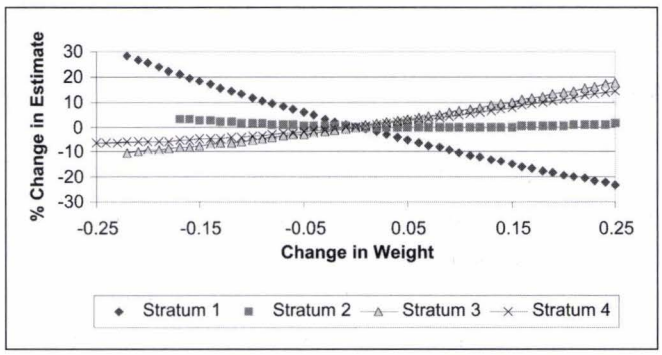


Figure 2.—*Percent change in variance of estimate of total volume when stratum weights are changed for four strata in Androscoggin County, Maine.*

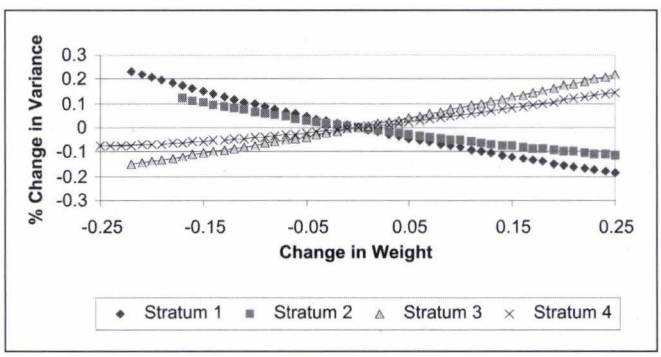


Figure 3.—*Change in percent sampling error for total volume when stratum weights are changed for four strata in Androscoggin County, Maine.*

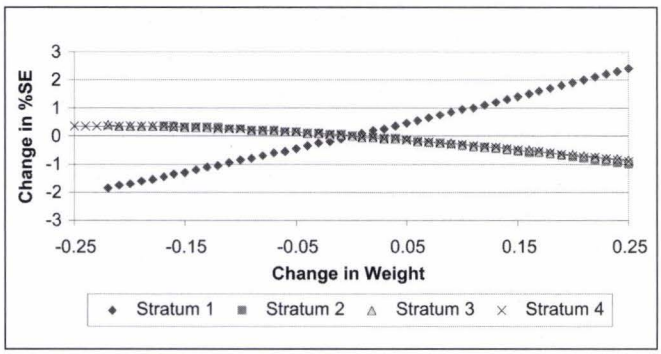
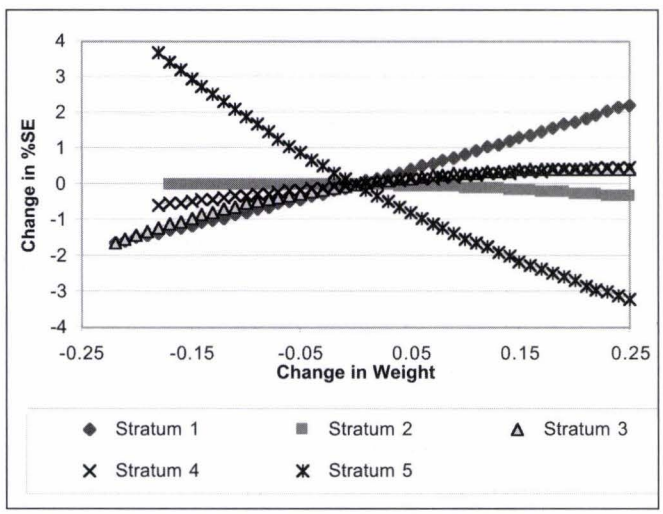


Figure 4.—*Change in percent sampling error for total volume when stratum weights are changed for five strata in Androscoggin County, Maine.*



associated with the weight changes. These changes in volume estimates and variances affect the percent sampling error (%SE). For example, figure 3 shows how the reduction in variance associated with decreasing weight for stratum 1 is exceeded by the accompanying decrease in the estimate with a net result of a smaller %SE. Strata 2 through 4 show surprisingly similar change in %SE trends. For these strata, a decrease in weight increases the %SE, as the increase in variance exceeds that of the estimate. When weight is added to these strata, the reduction in variance occurs more rapidly than the change in the estimate and %SE decreases.

Determining the number of strata also can be affected by classification algorithm, image selection, and other factors. For FIA purposes, Androscoggin County was divided into four strata (see table 1). For this study, the number of strata was altered to determine how different numbers of strata might affect estimates of volume. Where four strata were reduced to three, changes in estimates and sampling errors were similar to the four-strata configuration. However, the sampling error was reduced from 13.8 percent to 13.2 percent (when the original set of stratum weights was used). This was due to a slight increase (0.6 percent) in the estimate and a decrease (7.6 percent) in the variance. The reduction in variance was unexpected, as reducing the number of strata generally increases variance (Cochran 1977). The result presented here occurred for several reasons. First, the two strata that were combined had similar means, which leads to the combined data providing additional observations that clustered about a similar value. Additionally, these strata have relatively few observations due to the small county size. The combination of strata increases

sample size, which has a notable effect on the variance. This phenomenon would likely not be observed when computing estimates for larger areas or where sampling intensity is higher.

When the number of strata was increased to five, the changes in estimates of volume for a given change in stratum weight were similar in magnitude to those for the three- and four-strata analyses. However, the relationship between the rates of change differed. This is apparent in the trends for change in %SE (fig. 4), where decreases are now evident when weight is decreased for strata 3 and 4. Altering of weights in the newly created stratum 5 has the greatest effect on %SE. For this stratum, decreases in weight result in increases in %SE. The %SE based on original weights for the five-strata analysis was much lower (11.0 percent) than the %SE for the original four strata (13.8 percent). The improved sampling error resulted from the decrease in variance of 37.1 percent, which resulted from creating the addition stratum.

In each of the analyses presented, differences between the original estimate and those where stratum weights were altered were not significantly different at the 95-percent confidence level for stratum weight changes less than 0.20. Where the change in stratum weight exceeded 0.20, differences between estimates were significant in some cases.

Discussion

We can determine mathematically the magnitude and direction of change in volume estimates given a particular re-weighting of strata if the mean plot volumes of the strata are known. For example, when changes in stratum weights are between only two strata, the change in the estimate is proportional to the difference between the mean plot volumes of these strata. However, when weights are altered in more than two strata, the distribution of the change in weights also affects the outcome. Rates of change are determined by how far the stratum mean deviates from the overall sample mean. The further the stratum mean is from the sample mean, the greater the rate of change. The direction of change depends on whether the stratum mean is smaller or larger than the sample mean. The estimate of volume decreases when weight is added to strata whose mean is smaller than the overall mean. Similarly, the estimate increases as weight increases for strata with means greater than the overall mean.

It is not as clear how changing stratum weights alters variance estimates as strata usually have different mean volume estimates, while the number of plots and variances across strata may or may not be similar. The amount and direction of change in variance depends on the magnitude of differences in the variances of the individual strata. More importantly, changes in both the estimate and the variance can have a notable effect on the %SE. In instances where the mean plot volume for the strata is close to the overall mean, the change in variance drives the change in %SE. For strata with means that differ greatly from the overall mean, the change in the volume estimate is more influential.

As was illustrated for the five-strata analysis, a notable reduction in %SE is attainable. Although this reduction was essentially manufactured by the method by which the stratum was separated, it does show that certain relationships among the strata can help reduce sampling error. This does not mean that having more strata is always better as %SE also was reduced when changing from four strata to three. The key is to minimize the variances of the individual strata to the extent possible. Thus, if the merging of two strata results in a variance that is less than the sum of the original two strata, %SE will decrease. Likewise, breaking a single stratum into smaller strata can be beneficial if the sums of the variances for the smaller strata are less than the variance of the single stratum. The ability to reduce these variances will depend on both the spread of the data and number of observations. In this exercise, we used the plot data to increase the number of strata and minimize the variance within each stratum. In practice, this would not be possible as the 5 x 5 moving window summary does not give actual values for sample-plot data. Research is needed to create methodologies that provide classifications that are highly correlated with sample-plot data.

Conclusion

Stratification is an effective tool for improving the precision of FIA volume estimates. In many instances, stratified estimation procedures produce sampling errors that meet or exceed guidelines

for estimates of area and volume. However, in some situations it is difficult to obtain sufficiently low sampling errors (e.g., estimation for small areas). This may preclude drawing meaningful conclusions. Our research has shown that it is possible to improve sampling errors by refining stratification techniques. A stratification method that optimizes the number of strata and associated variances might be an efficient and effective way to obtain meaningful estimates where the number of measurement plots is small. This approach would likely be far less expensive and time consuming than increasing sampling intensity.

Literature Cited

Cochran, W.G. 1977. Sampling techniques. 3d ed. New York, NY: John Wiley and Sons. 428 p.

Scott, C.T. 1981. Northeastern forest survey revised cubic-foot volume equations. Res. Note NE-304. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 3 p.

U.S. Census Bureau. 2001. TIGER/Line Files, Census 2000. Washington, DC: U.S. Department of Commerce, Census Bureau. Available: http://www.census.gov/geo/www/tlmetadata/tl2kmeta.txt [12 October 2001].

U.S. Department of Agriculture. 2002. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.6. Washington, DC: U.S. Department of Agriculture, Forest Service. Internal report.

Vogelmann, J.E.; Sohl, T.L.; Campbell, P.V.; Shaw, D.M. 1998. Regional land cover characterization using Landsat thematic mapper data and ancillary data sources. Environmental Monitoring and Assessment. 51: 415–428.

Sensitivity Analysis of Down Woody Material Data Processing Routines

Christopher W. Woodall[1] and Duncan C. Lutes[2]

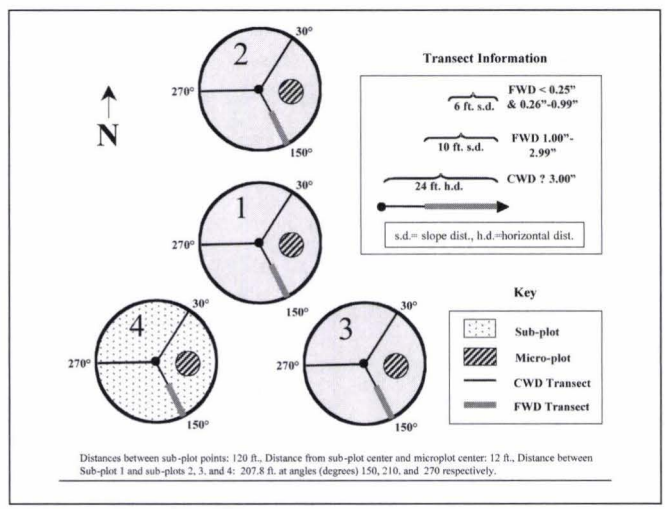
Abstract.—Weight per unit area (load) estimates of Down Woody Material (DWM) are the most common requests by users of the USDA Forest Service's Forest Inventory and Analysis (FIA) program's DWM inventory. Estimating of DWM loads requires the uniform compilation of DWM transect data for the entire United States. DWM weights may vary by species, level of decay, woody material type, and size. Additionally, weight estimates may vary by compilation constants and methods. To better facilitate DWM compilation routines, the effect of the variation in fuel processing routine constants and measurement error of variables on the resulting DWM load estimates was examined. Sensitivity analysis indicated that some compilation constants and measurement variables disproportionately influenced load estimates of DWM. More accurate and efficient estimates of DWM components may be acquired by identifying compilation constants and measurement variables that are the largest sources of variation in weight estimates.

Down woody material (DWM) is the dead material on the forest floor in various stages of decay. Down woody components estimated by the Forest Inventory and Analysis (FIA) program are coarse woody, fine woody, litter, herb/shrubs, slash, duff, and fuelbed depth. As defined by the FIA program, coarse woody debris (CWD) is down logs ≥ 3 inches in transect diameter and ≥ 3-feet long. Transect diameter is the diameter of a down woody piece at the point of intersection with a sampling transect. Fine woody debris (FWD) is down woody materials with a transect diameter less than 3 inches. Slash piles are collections of down coarse woody debris, whether from logging or natural disturbances. Shrubs are defined as non-tree woody vegetation. Herbs are non-woody herbaceous plants, but also include ferns, moss, lichens, sedges, and grasses. Litter is dead plant material on the forest floor excluding CWD, FWD, and duff. Duff is decomposed plant material beneath the litter layer with no identifiable plant parts (i.e., stems and leaves) included.

DWM is sampled during a specific phase of FIA's multiscale inventory sampling design. The FIA sampling design consists of three phases. The first phase superimposes a hexagonal grid across forest/nonforest maps of the United States. Each hexagon (approximately 6,000 acres in area) contains one sample point. If the sample point falls on a forested area, a field crew will visit the location and establish a permanent sample plot (a phase 2 plot). All phase 2 plots are measured for tree and site attributes. Phase 2 plots are subsampled (phase 3) (approximately 1/16 of all phase 2 plots, 96,000 acres) for indicators of forest health such as DWM. Due to the low sampling intensity and application of data to address multitudes of regional issues, the DWM inventory is appropriately termed the DWM Indicator.

The sampling design of the DWM Indicator is a combination of planar intersect, point, and microplot sampling (U.S. Department of Agriculture, Forest Service 2002) (fig. 1). CWD

Figure 1.—*The Forest Inventory and Analysis Program's (USDA Forest Service) Down Woody Materials sampling design.*



[1] Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108; Phone: 651–649–5141; e-mail: cwoodall@fs.fed.us.

[2] Research Forester, Systems for Environmental Management, P.O. Box 8868, Missoula, MT 59807.

and FWD are sampled on transects radiating from each FIA subplot center. Information collected for every CWD piece intersected on each of three, 24-foot transects on each FIA subplot is transect diameter, length, small-end diameter, large-end diameter, decay class, species, evidence of fire, and presence of cavities (fig. 1). FWD with a transect diameter of 0 to 0.99 inches (1-hr and 10-hr) is tallied on a 6-foot slope-distance transect (one transect per FIA subplot) (fig. 1). FWD with a transect diameter of 1.00 to 2.99 inches is tallied on a 10-foot slope-distance transect (one transect per FIA subplot) (fig. 1). The duff and litter are sampled using a point estimate of depth at a 24-foot slope-distance along each CWD transect (for a total of 12 sample points). The shrub and herb fuel complex is sampled on the phase 2 microplot (6.8-foot radius) (fig. 1). The percentage cover (10 percent classes) and total height of dead and live shrubs/herbs (including grasses) is estimated. Slash piles with centers that are within 24 feet of any subplot center are sampled, using methodology developed by Hardy (1996). The shape of each slash pile is classified into four slash pile shapes. Based on the pile shape classification, appropriate dimensions of the slash pile are measured along with an estimate of pile density.

DWM inventory field data are organized into seven database tables reflecting the various components estimated by the DWM sampling design: CWD, FWD, microplot, transect information, plot information, duff/litter, and slash piles. Although invaluable to numerous research initiatives, the seven tables of DWM data need to be processed to produce plot estimates of DWM components. Just as basal area/acre estimates are determined for phase 2 plots, users of FIA data desire weight per unit area (load) estimates of DWM. The seven tables of DWM data may be processed in many ways to obtain per acre estimates of DWM components. For a more complete guide to FIA's DWM sampling design, please refer to U.S. Department of Agriculture, Forest Service (2002) or field guides from any of the other regional FIA units.

The goal of our study was to ascertain through sensitivity analysis the impact of variations in data processing techniques and measurement errors on the final load estimates of DWM components. Sensitivity analysis results were used both to identify critical parts of DWM data processing algorithms and to manage the quality analysis and control of the DWM inventory.

Processing Algorithms

Brown (1974) originally summarized many of the sampling protocols adopted by the DWM Indicator. In addition to the sampling design, Brown (1974) provided numerous load-processing models for DWM components. Although numerous DWM data processing algorithms are possible, the basic models of Brown (1974) and slash pile models by Hardy (1996) were used in the sensitivity analysis.

FWD data were processed using the following constants and measurement variables: unit's conversion constant (convert sampling measurement units to tons/acre), number of tallied FWD pieces, quadratic mean diameter (QMD) of the appropriate FWD size class, specific gravity, non-horizontal angle correction factor (corrects for DWM pieces nonperpendicular to transect line), slope, and transect total length (Brown 1974). CWD data were processed using the following constants and measurement variables: unit's conversion constant, sum squared CWD transect diameters, specific gravity, nonhorizontal angle correction, slope, and total transect length (Brown 1974). Litter and duff data were processed using mean depths, specific gravity, and a unit's conversion constant (Brown 1974). Slash pile data were processed using pile volume based on pile shape equations (Hardy 1996), specific gravity, and slash packing ratio (amount of wood occupying volume defined by pile dimensions).

Methods

The DWM inventory data from its first year of implementation, 2001, were used in this study. Over 900 plots were used in the analysis from 32 States distributed across the U.S.

The effect of the variation in DWM processing routine constants and the measurement error of variables on the resulting DWM load estimates was examined using sensitivity analysis. The effect of 5-, 10-, and 15-percent variation in various selected DWM measurement variables and processing constants was evaluated in terms of final load estimates. Constants used in the sensitivity analysis include CWD specific gravity, CWD decay rate, litter specific gravity, FWD QMD, duff specific gravity. Measurement variables used in sensitivity analysis

include FWD 0-0.25 inch tally counts, litter depth, CWD transect lengths, FWD transect lengths, slash pile heights, slash pile packing ratio, CWD transect diameter, and duff depth.

Results/Discussion

Sensitivity analysis of DWM processing constants indicated that variation in constant values had disproportionate effects on total load estimates (fig. 2). A 5 percent increase in the specific gravity of a plot's CWD pieces resulted in approximately a 1 percent increase in the plot's total DWM load estimate, whereas a 5 percent increase in the specific gravity of duff resulted in nearly a 2.5 percent increase in the plot's DWM load estimate. Obvious from sensitivity analysis, constants such as the specific gravity of duff and the QMD of FWD may have the greatest effect on resulting determinations of plot DWM load estimates (fig. 2). Because duff usually has a far greater specific gravity than litter, variations in its estimate can greatly impact on the total plot DWM load estimate. When processing DWM inventory data, special attention should be given to selecting constants that influence total DWM loading estimates the most.

Sensitivity analysis of DWM measurement variables essentially is a review of the effect of measurement error on total DWM plot load estimates. Our analysis showed a disproportionate effect of variation in certain measurements on resulting variation in plot estimates. FWD tally counts, litter depth, and CWD transect lengths had a minor impact on plot totals: a 15 percent increase in their values resulted in less than a 3 percent variation in DWM plot estimates (fig. 3). For variables such as slash packing ratio, CWD transect diameter, and duff depth, a 15 percent increase in their associated values resulted in a greater than 5 percent variation in DWM plot estimate (fig. 3). Since duff and CWD components typically contain substantial tonnage, variations in their processing routines and/or measurement errors may have the greatest effect on resultant plot DWM estimates. Obviously, these sensitivity analysis results would not necessarily apply to individual DWM components such as FWD.

Sensitivity analysis results indicate that the selection of any DWM processing routine may initially hinge on which DWM components contribute the most to overall DWM plot estimates. Since constants used to determine CWD and duff tonnage estimates might greatly affect output, those processing routines should be scrutinized first. For instance, DWM analysts should concentrate more effort on which duff specific gravity is selected for a plot than on what CWD specific gravity is selected for a certain CWD piece on a plot. Results from the sensitivity analysis of DWM variables (i.e., analysis of measurement error) have implications for DWM data quality assessment/quality control (QA/QC). Currently the QA/QC tolerance for measurement of duff depth is ± 0.5 inches. For the 2001 field season a 0.5-inch variation in duff depth would on average be a 20 percent measurement error. Based on this study's sensitivity analysis, 20 percent measurement error in duff depth would result in a 9 percent error in plot DWM esti-

Figure 2.—*Effect of 5, 10, 15% variation in various down woody material processing constants on total per acre tonnage estimates.*

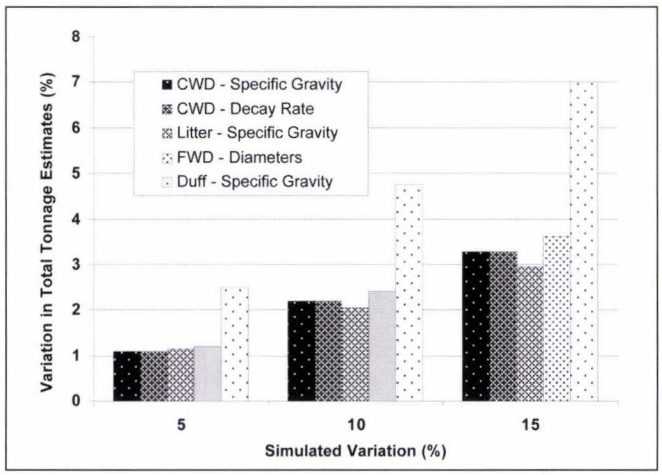
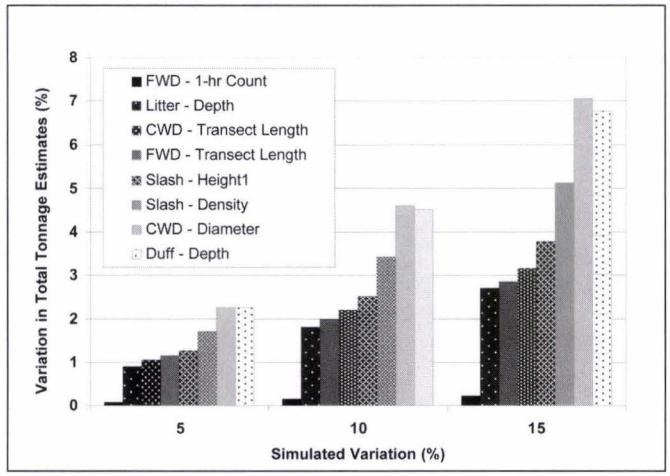


Figure 3.—*Effect of 5, 10, 15% variation in down woody material measurement variables on total per acre tonnage estimates.*



mates of total load. The QA/QC tolerance of FWD 0-0.25-inch-class tally counts is ± 20 percent. Based on sensitivity analysis, a 20 percent measurement error of this FWD component will typically result in a 0.5 percent error in plot level estimates of total load. If users of FIA data are more interested in plot estimates of DWM components, QA/QC efforts might be better derived from a "top down" approach where plot estimate variation drives measurement error tolerances. If FIA customers desire total tonnage/acre estimates, reductions in measurement errors that greatly affect those load estimates should be undertaken rather than arbitrarily setting measurement error tolerances.

Conclusions

The DWM Indicator of the FIA program provides the first nationally consistent inventory of DWM components. The DWM data sets quantify a structurally diverse component of forest ecosystems. Although FIA data are invaluable as a research database in its "unprocessed" form, numerous users desire plot estimates of DWM. These estimates fulfill data requirements of numerous forest research initiatives ranging from the fire sciences to carbon budget accounting. There are many methods for processing DWM data and many sources of measurement error. The sensitivity results of this study may help refine the debate among those that process DWM data and those that manage the quality of the DWM data. Some processing constants and measurement variables (duff depths, specific gravity, CWD diameters, and slash pile dimensions) had disproportionately greater effect on total plot load estimates than other variables (CWD decay rates/classes/transect lengths, litter depth/specific gravity, and FWD counts), which had a minimal effect. The processing and QA/QC of the DWM Indicator may be further refined with a better understanding of data outputs desired by FIA constituencies, more holistic comprehension of how all DWM components interact during data processing, and QA/QC guidelines determined by analysis of actual field data.

Literature Cited

Brown, J.K. 1974. Handbook for inventorying downed woody material. Gen. Tech. Rep. INT-16. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Hardy, C.C. 1996. Guidelines for estimating volume, biomass, and smoke production for piled slash. Gen. Tech. Rep. PNW-364. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

U.S. Department of Agriculture, Forest Service. 2002. Forest inventory and analysis national core field guide; volume II: field data collection procedures for phase 3 plots, version 1.6, section 14: 265–298.

Simulating Silvicultural Treatments Using FIA Data

Christopher W. Woodall[1] and Carl E. Fiedler[2]

Abstract.—Potential uses of the Forest Inventory and Analysis Database (FIADB) extend far beyond descriptions and summaries of current forest resources. Silvicultural treatments, although typically conducted at the stand level, may be simulated using the FIADB for predicting future forest conditions and resources at broader scales. In this study, silvicultural prescription methodologies were simulated using FIA inventory plots from Montana that rated high for fire hazard. Database operations were used to couple the FIADB with the silvicultural prescriptions process, allowing successful simulation of silvicultural treatments. Cut- and leave-tree tables were created for each FIA sample plot using computer "marking" algorithms, allowing estimation of current and future forest attributes (volume, growth, wildfire hazard, and treatment costs). Simple database operations can be used to mimic complex silvicultural prescriptions using FIA inventory data from major ownerships, States, or regions, allowing evaluation of treatment effects and future forest conditions.

Forest Inventory and Analysis (FIA) data have traditionally been used to summarize current forest conditions. Forest resource assessments for States and regions, for example, have been invaluable sources of information for forest managers and decisionmakers alike. However, there is a growing need not only for current forest assessments, but also for evaluation of various management alternatives. In other words, given the current status of forests as inventoried by FIA, what would be the effect of a specific management action applied to said forest acreage?

Given the technological advances in database management and forest ecosystem modeling, a large portion of FIA analysis and reporting in the future may involve simulating management actions on current forest conditions. Just as FIA was founded over 75 years ago to answer the question, "how much forest," more timely and complex questions now arise, such as "given current forest conditions, what might be the effect of specific management treatments?"

One of the most immediate and controversial forest management issues nationally involves the fire hazard problem in the West, and the kinds and costs of treatments being proposed to address it. For example, what forest types are most vulnerable to crown fires? What silvicultural treatments are most effective for reducing stand-level fire hazard, and how much do they cost? At a statewide or regional level, what is the potential contribution of a strategic hazard reduction program to the Nation's wood supply, and how much might it cost? The consistent and comprehensive data collected by the FIA inventory provide a uniform and objective basis for addressing these questions. Toward this end, we developed a methodology for simulating silvicultural treatments in the FIADB environment and evaluating their effectiveness for reducing hazard in the dry, low elevation forests of Montana.

Silvicultural Prescription Process

To mimic hazard reduction treatments using the FIADB, the silvicultural treatment process needs to be reduced to its elemental steps and incorporated into database logic. The silvicultural prescription process involves five distinct steps: inventory assessment, diagnosis, prescription development, prescription implementation, and evaluation/monitoring.

Inventory Assessment.—Inventory is quantified or translated in terms of stand density, structure, and species composition for use in the silvicultural prescription process. Density is expressed in terms of trees per acre, basal area per acre, or

[1] Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108; Phone: 651–649–5141; e-mail: cwoodall@fs.fed.us.
[2] Associate Professor, School of Forestry, University of Montana, MT 59802.

Stand Density Index (SDI) (Reineke 1933), with location, site, and species used to determine break points between high-, medium-, and low-density classes. Structure is described in terms of trees, basal area, or SDI per acre by diameter class, with species and site criteria used to classify conditions into one-, two-, or multi-storied structures. Species composition is typically expressed in terms of percent basal area representation by species, which is used in conjunction with habitat type criteria (Pfister *et al.* 1977) to classify forest types. The FIADB is a robust source for all of these data, providing the necessary inputs for the silvicultural prescription process.

Diagnosis.—The diagnosis of individual forest stands is based on the inventory. Inventory data are interpreted in terms of location and ecological context to assess current conditions versus those desired to achieve the management objective. For the Montana FIA inventory, each plot may be treated as an individual stand, where its current stand density, structure, and species composition can be used to diagnose fire hazard and priority for fuel-reduction treatment.

Prescription Development.—For every stand evaluated in the prescription process, a target future stand condition is developed. A "target" stand specifies the density, structure, and species composition that best contribute to meeting the management objectives. In this study, for example, target conditions to achieve low fire hazard would include low to moderate stand density, one-, two-, or multi-storied structures dominated by large trees, predominantly ponderosa pine. Current conditions are diagnosed from the inventory, and compared to the target conditions. To the extent that existing conditions differ from and are not trending toward the target conditions over time, some form of management intervention is indicated. This leads to the third step in the process – prescription development. In this step, one or more treatments are designed or selected for manipulating existing forest conditions to create the desired conditions, either immediately or over time.

Implementation.—The fourth step of the prescription process is implementation, which involves on-the-ground management activities ranging from silvicultural cutting to treatment or removal of activity fuels. If silvicultural treatments are simulated for FIA inventoried forest acreage, then implementation involves database manipulation of tree records in accordance with prescription guidelines.

Evaluation/Monitoring.—The final stage of the prescription process involves evaluation and monitoring of stand conditions following prescription implementation. Post-treatment stand conditions are typically assessed to determine if the prescribed treatments had the desired effects. In an FIA inventory context, this step in the prescription process involves modeling and estimating stand attributes such as fire hazard, treatment costs/revenues, and future growth responses.

Simulating the Prescription Process

The 1999–2000 periodic FIA inventory for Montana (USDA INT 1999a, 1999b) was used in this simulation of silvicultural treatments to reduce crown fire hazard. Only dry, low- to mid-elevation forests were included because these forests show the greatest departure from historically sustainable conditions (Fiedler *et al.* 2003). All silvicultural prescription process steps were followed to simulate treatments within the FIADB:

1. Inventory.—All FIADB inventory plots were screened to select those that qualified as ponderosa pine, Douglas-fir, or dry lower mixed conifer forest types, based on species composition and habitat type.
2. Diagnosis.—Conditions that contribute to high fire hazard (high stand density, complex structure, composition of late-seral species) were identified.
3. Prescription development.—Characteristics of a target stand with low fire hazard (i.e., low to moderate density, structure predominated by large ponderosa pine) were incorporated into the comprehensive prescription.
4. Prescription implementation.—Two prescriptions (thin-from-below and comprehensive) were proposed to reduce fire hazard in subject FIA plots (stands).
5. Evaluation and monitoring.—Stands (pre- and post-treatment) were evaluated for crown fire hazard using the Fuel and Fire Extension (Scott and Reinhardt 2001) of the Forest Vegetation Simulator (Beukema *et al.* 2000, Crookston and Havis 2002, Wykoff *et al.* 1982). Harvest

costs, product values, and slash treatment costs were estimated using harvest cost models (Keegan *et al.* 2001), databases (Bureau of Business and Economic Research 2001), and data collected from land management agencies and the private sector, respectively.

The thin-from-below prescription (TB) removes all trees up to 9 inches diameter, followed by treatment of activity fuels. The comprehensive prescription (COMP) marks 40-50 square feet per acre of mostly large serals to leave. Then it uses low thinning to greatly reduce ladder fuels, and improvement/selection cutting to reduce density and remove undesirable trees in the mid/upper canopy, followed by treatment of activity fuels (Fiedler 2000; Fiedler *et al.* 1999, 2001).

The two treatment prescriptions were applied to selected FIA plots through development of a "marking" algorithm. The algorithm feeds the inventory database table through a decision matrix (silvicultural prescription logic) where each individual tree record is sent either to a leave table or cut table, depending on the tree's size and species and on characteristics of other trees in the stand. Leave tables are used to summarize post-treatment stand conditions and model fire hazard. Cut tables are used to determine the cost/revenue of implementing the silvicultural cuttings and treating the slash.

The marking algorithm decision matrix varies by prescription. For the TB prescription, trees sorted to the leave table were greater than or equal to 9 inches in d.b.h. without considering residual stand density, structure, or species composition. For the COMP prescription, trees sorted to the leave table were based on an iterative process of selection preferences based on density, structure, and species composition. Density of the reserve stand was set at 45 square feet per acre for this simulation. Species preference was set in order of ponderosa pine, western larch, lodgepole pine, and Douglas-fir. Desired stand structure was set on the basis of a target basal area per acre by 4-inch diameter classes: >20" (20 ft^2/ac); 16-20" (10 ft^2/ac); 12-16" (7 ft^2/ac); 8-12" (5 ft^2/ac); 4-8" (2.5 ft^2/ac); and <4.0" (0.5 ft^2/ac). If insufficient basal area density was present in any given diameter class, a second set of logic rules was used to "borrow" basal area from other d.b.h. classes as needed to reach the density target of 45 square feet per acre, while still approximating the desired structure.

Simulation Output

The methods used in this study successfully simulated the silvicultural prescription process, even the complex comprehensive prescription that integrated several silvicultural cutting treatments. Using each FIADB plot as an individual stand unit and corresponding tree records as constituents of a hypothetical "stand table," allowed simulation of forest management activities at a statewide scale. Sorting the FIADB tree records through a "marking" algorithm into cut and leave tables enabled modeling post-treatment stand conditions and crown fire hazard (fig. 1), as well as net revenues/costs associated with alternative hazard reduction prescriptions (figs. 2 and 3).

Output of the treatment simulations showed that the COMP treatment reduced fire hazard for most treated stands. Crowning index, which is the windspeed necessary to maintain

Figure 1.—*Distribution of acres by crowning index for existing high-hazard forest conditions, and after thin-from-below and comprehensive restoration treatments (adapted from Fiedler* et al. *2004).*

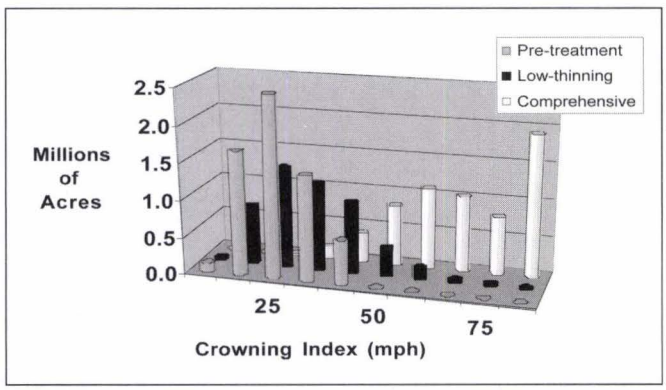


Figure 2.—*Distribution of acres by net revenue for the comprehensive restoration treatment (adapted from Fiedler* et al. *2004).*

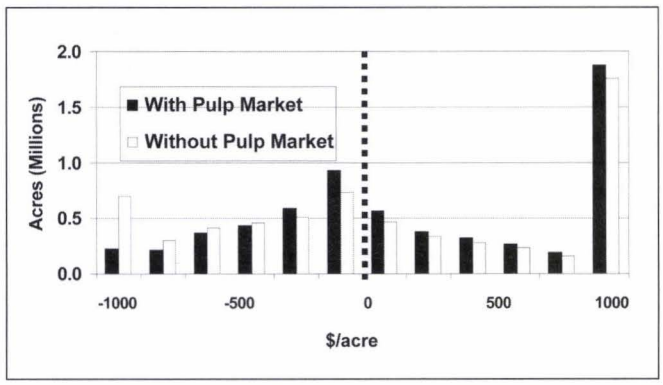
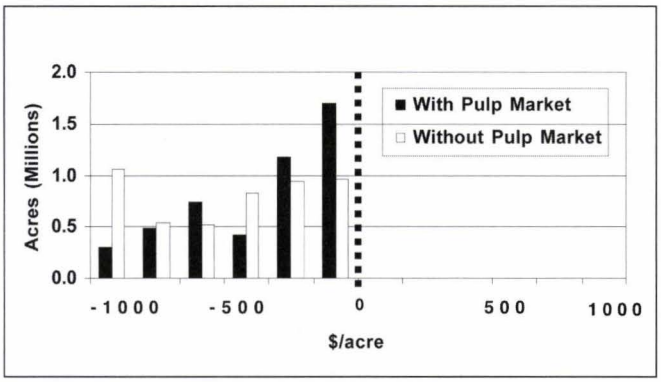


Figure 3.—*Distribution of acres by net revenue for the thin-from-below treatment (adapted from Fiedler* et al. *2004).*



a crown fire once it has reached the main canopy, increased dramatically following application of this treatment (fig. 1). In contrast, TB treatment had only a slight effect on crown fire hazard, with little change in crowning index from pre-treatment levels (fig. 1). By using the cut-tree stand table to determine treatment costs/revenues, it was evident that a large portion (approximately half) of the stands receiving the COMP treatment produced net revenues (fig. 2), while all stands receiving the TB treatment required out-of-pocket expenditures (fig. 3). A complete account of simulation results is presented in Fiedler *et al.* (2004).

Conclusions

Silvicultural prescriptions can be simulated for large areas (e.g., major ownerships, States) by using basic database management procedures to manipulate tree records in the FIADB. The database procedures involve use of a marking algorithm to sort individual tree records into separate leave and cut database tables. These leave- and cut-tree tables allow modeling of post-treatment fire hazard and the revenues/costs associated with treatment activities. This approach also allows managers to design hazard reduction treatments that are both effective and cost-efficient. In addition, it can help policymakers evaluate hazard reduction treatments for our Nation's forests before they are ever implemented on the ground.

Acknowledgment

We thank the Joint Fire Sciences Program for funding the study *A Strategic Assessment of Fire Hazard in Montana* on which this paper is based.

Literature Cited

Beukema, S.J.; Reinhardt, E.; Greenough, J.A.; *et al.* 2000. Fire and fuels extension: model description. Working draft prepared by ESSA Technologies Ltd., Vancouver, B.C. for U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Moscow, ID. 58 p.

Bureau of Business and Economic Research. 2001. The log price reporting system. Missoula, MT: The University of Montana.

Crookston, N.L.; Havis, R.N., eds. 2002. In: Proceedings, 2nd forest vegetation simulator conference; 2002 February 12–14; Fort Collins, CO. Proc. RMRS-P-25. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Fiedler, C.E.; Keegan, C.E.; Arno, S.F.; Wichman, D.P. 1999. Product and economic implications of ecosystem restoration. Forest Products Journal. 49: 19–23.

Fiedler, C.E. 2000. Understanding the ecosystem: its parts and processes - silvicultural treatments. In: Proceedings, the Bitterroot ecosystem management research project: what we have learned. RMRS-P-17. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 19–20.

Fiedler, C.E.; Arno, S.F.; Keegan. C.E.; Blatner, K.A. 2001. Overcoming America's wood deficit: an overlooked option. BioScience. 51: 53–58.

Fiedler, C.E.; Keegan, C.E., III; Woodall, C.W.; Morgan, T.A. 2004. A strategic assessment of crown fire hazard in Montana: potential effectiveness and costs of hazard reduction treatments. Gen. Tech. Rep. PNW-GTR-622. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 48 p.

Keegan, C.E.; Niccolucci, M.J.; Fiedler, C.E.; Jones, J.G.; Regel, R.W. 2001. Harvest cost collection approaches and associated equations for restoration treatments on national forests. Forest Products Journal. 52: 96–99.

Pfister, R.D.; Kovalchik, B.L.; Arno, S.F.; Presby, R.C. 1977. Forest habitat types of Montana. Gen. Tech. Rep. INT-34. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 174 p.

Reineke, L.H. 1933. Perfecting a stand-density index for even-aged stands. Journal of Agricultural Research. 46: 627–638.

Scott, J.H.; Reinhardt, E.D. 2001. Assessing crown fire potential by linking models of surface and crown fire behavior. Res. Pap. RMRS-RP-29. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 59 p.

U.S. Department of Agriculture. 1999a. Montana forest survey field procedures. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

U.S. Department of Agriculture. 1999b. Interior West forest resource inventory field procedures. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Wykoff, W.R.; Crookston, N.L.; Stage, A.R. 1982. User's guide to the stand prognosis model. Gen. Tech. Rep. INT-133. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 14 p.

Spatial Information Needs on the Fishlake National Forest: Can FIA Help?

Robert B. Campbell, Jr.[1], and Renee A. O'Brien[2]

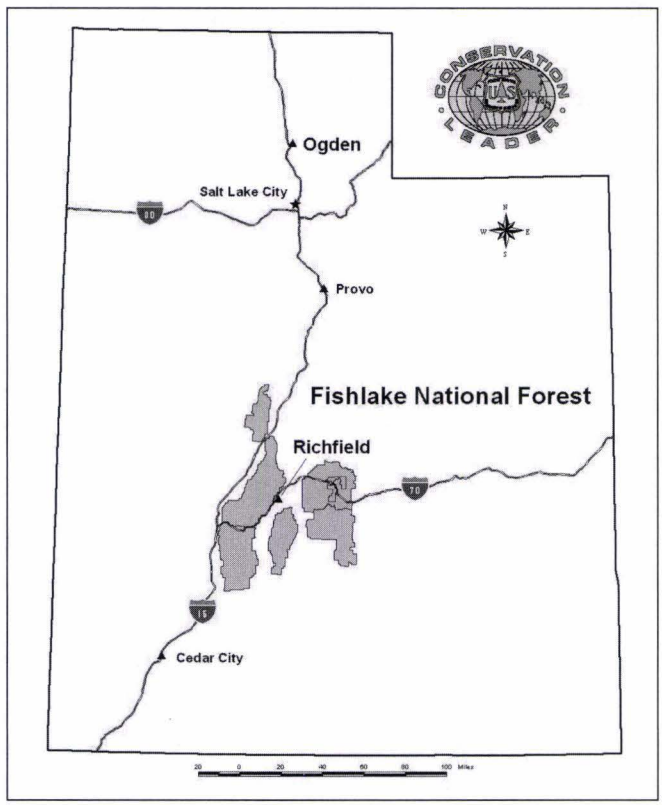
Abstract.—National forest staff members are frequently challenged to make assessments with existing information. They rarely have the time or resources to go to the field to gather new data specific to the question at hand. Forest Inventory and Analysis (FIA) data have proved useful in the past, but there is an increasing need for spatial depictions of forest resources to address management and planning issues. For example, maps are needed to assess healthy stands, suitable wildlife habitat, marketable harvest areas, desired future conditions, and historical distribution of forest types. The success of FIA-generated map products hinges on good communication with map users throughout the mapmaking process, adequate development and accuracy assessment, ecological integration, and rigorous field testing.

This paper introduces the Fishlake National Forest (FNF) and describes the collaboration that occurs between the Rocky Mountain Research Station's (RMRS) Interior West Forest Inventory and Analysis (IWFIA) unit in Ogden, Utah, and FNF. This paper also provides the contextual backdrop for multiple papers presented at this conference that report various projects underway on the FNF (Edwards *et al.* 2004, Frescino and Moisen 2004, Schultz, R.J. 2004, Terletzky and Frescino 2004). The FNF benefited immensely from interactions with the RMRS-IWFIA staff in Ogden.

The Forest needs mapped information, and FIA products that spatially refine data would benefit the FNF. This paper presents a background discussion on the goals of the Forest to sustain biodiversity and maintain properly functioning ecosystems. It follows with a brief historical background of the utility of past IWFIA products and the coordination between the Forest and IWFIA for meeting the needs of the Forest. It ends with a discussion of the current situation in the Forest, and again, the need and desire for IWFIA spatial products. The paper is as much about the *process* as it is about the *products*. Thus, this paper describes the synergy that results from the collaboration of Forest managers and specialists with Station scientists and researchers.

The FNF occupies about 1.5 million acres in south-central Utah (fig. 1). The Supervisor's office is in Richfield on I-70, 40 miles east of the western terminus of I-70 with I-15. The Forest features incredible landscape and biological diversity. Elevations range from 5,000 feet to over 12,000 feet. The highest point lies in the southwestern part of the Forest; Delano Peak in the Tushar Mountains stands at 12,169 feet. In addition to ragged peaks, sweeping high elevation plateaus are blanketed with mixed-conifer and aspen forests. The Forest's eastern edge is bounded by the arid, rugged terrain of Capitol Reef

Figure 1.—*Vicinity map of Utah and the Fishlake National Forest.*



[1] Ecologist, U.S. Department of Agriculture, Forest Service, Fishlake National Forest, 115 East 900 North, Richfield, UT 84701.

[2] Lead Ecologist, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, UT 84401–2394.

National Park and the San Rafael Swell. Annual precipitation on the Forest varies from less than 8 inches to more than 40 inches. Recent assessments focused on Monroe Mountain, the Beaver River Watershed, and the Fishlake Basin/Sevenmile Creek in the upper Fremont River drainage. Fish Lake, at 8,800 feet and about 5 miles long by 1 mile wide, is deep, cold, and considered by many to be Utah's "crown jewel." Wilderness areas, either existing or proposed, do not occur on the Forest. Resource themes important to the Forest's landscapes include fuels, timber, and wildlife. Also, the FNF is in the initial revision phase of land and resource management planning.

Definitions and Concepts

Biodiversity and Properly Functioning Condition

Biological diversity is often described in terms of composition, structure, and function. Composition is described by the numbers and kinds of plants and animals. Structure relates to the sizes, shapes, and/or ages of the plants and animals. Function (or process) has to do with what happens in the ecosystem. For example, disturbance regimes like fire, flood, or windthrow are all types of functions and ecosystem processes. Also, function includes the contribution each plant and animal species provides to the ecosystem.

The Forest Service's Intermountain Region began a process in 1996 that expanded the Bureau of Land Management's concept of proper function condition in riparian areas to the properly functioning condition (PFC) of the major upland vegetation cover types (USDA Forest Service 1998, 2000a). This PFC approach provided an ecological basis for the rapid assessment of general conditions of sustainability on large landscapes. Properly functioning condition is defined with this statement (USDA Forest Service 1998, 2000a, 2000b):

> Ecosystems at any temporal or spatial scale are in properly functioning condition when they are dynamic and resilient to perturbations to structure, composition, and processes of their biological or physical components.

Because that definition is fairly technical, Campbell and Bartos (2001) suggest another definition for use with general audiences (e.g., school classes or public meetings) that is less technical, yet attempts to convey the same meaning:

> Properly functioning condition exists when soil and water are conserved, and plants and animals can grow and reproduce and respond favorably to periodic disturbance.

PFC is not a single state in space or time. PFC includes a range of situations and conditions that allow for the full variation of composition and structure within the processes of sustainable functioning ecosystems for that specific major vegetative cover type.

Often our stakeholders, both internal and external, Forest employees, county commissioners and city leaders, Forest users (permittees, recreationists, loggers, summer home owners, etc.) and students (younger or older) really do not care what the statistical difference or R^2 is if they can not see the difference on the ground. Maps that display the elements of biodiversity, particularly composition and structure, would be useful to describe and explain concepts of biodiversity.

Major Vegetation Communities and Biodiversity Loss

Major vegetation cover types on the FNF include spruce/fir, aspen, mixed-conifer/aspen, ponderosa pine, curl-leaf mountain mahogany, Gambel oak, pinyon/juniper, and sagebrush/grass/forb. These cover types are fire adapted and have many traits that allow periodic fire to be a stimulant and a healthy process to sustain them.

Periodic fires maintain structural diversity within vegetation cover types. Cover type conversions, in the absence of fire, result in a loss of compositional biodiversity. The FNF now has an absence of historical fire regimes combined with substantial increases in ungulate use, both domestic and wildlife. With these changes in disturbance patterns, aspen is being replaced by spruce/fir; aspen is being replaced by sagebrush/grass/forb; and sagebrush grass forb is being replaced by pinyon/juniper. This results in loss of ecosystem function; biodiversity and sustainability are compromised. Maps of the structure and composition of major vegetation cover types would be especially useful to forest managers to demonstrate potential loss of function.

History, Applications, and Use of IWFIA Products

Second Generation Use of Data and Synergy

Forest specialists are often challenged to complete assessments with existing information. The second generation use of data is simply this: use previously gathered data to answer questions that were not conceived at the time the original data were collected. This results from collaboration that leads to synergy.

A landmark meeting involving IWFIA researchers and Intermountain Region employees was held in Ogden, Utah, in November 1995. Most of the six national forests in Utah were represented, and scientists from other Station research work units attended. The standard IWFIA products were displayed and discussed. Then the question was posed, what other products would be useful? It would be difficult to overstate the value of the *synergy* that began at that one meeting!

Resource Reports for Utah Forests

One idea expressed at the November 1995 meeting was the desire to have a brief report of the forest resources. The outgrowth of that idea expanded to a glossy, forest report for each of the six forests in Utah (O'Brien and Brown 1998, O'Brien and Collins 1997, O'Brien and Pope 1997, O'Brien and Tymcio 1997, O'Brien and Waters 1998, O'Brien and Woudenberg 1998).

Ron Sanden, FNF Forest Silviculturist (retired), described his experiences with and the value of the *Forest Resources of the Fishlake National Forest* (O'Brien and Waters 1998) (personal communication):

> I took copies of the Fishlake report to all of the meetings that I attended that first year or two. I love the report. It is the most valuable handout or pamphlet that I have used to explain the Fishlake's forest resources.

The collaboration and synergy continued as these reports were prepared. Additional input from FNF employees led to the development of a bar chart in the Fishlake report that displayed acres by age class for each of the forest types. Actually, the bar chart displayed the magnitude of structural diversity within each forest type. The information proved quite useful during various assessments completed on the FNF. Now we need this kind of information on structural diversity mapped and displayed spatially.

Initial Use of IWFIA Data

Another outgrowth of the November 1995 meeting was to query the IWFIA database in ways that demonstrated the magnitude of aspen decline in Utah. Aspen decline occurs when landscapes with aspen are outside of properly functioning condition. The results of these refined queries of the IWFIA database showed that the six forests in Utah have had nearly a 60-percent decline in aspen forest types from nearly 2.1 million acres to less than 0.9 million acres (Bartos and Campbell 1998a, 1998b). Again, managers would benefit greatly if this information were mapped.

The FNF completed a forestwide assessment of historical, existing, and desired vegetation conditions during 1997 and 1998 as a part of its Prescribed Natural Fire Plan (Jackson *et al.* 1998). (The current terminology is now Wildland Fire Use Plan.) The IWFIA data were extremely useful in developing the assumptions used to determine what the historical abundance of the major vegetation cover types had been. Forest Supervisor Rob Mrowka awarded Renee O'Brien a Certificate of Merit for special effort in collaboration and for displaying forest resources data that allowed interpretation of historical vegetation cover for the Fishlake National Forest's Prescribed Natural Fire Plan.

IWFIA data are also used to help people understand how much the Forest's landscapes have changed in the past 200 years. A table of acres in each stand-age class by forest type (O'Brien 1999) was derived from a consistent, uniform, FIA data set for all of Utah. The 50-year stand-age classes included nearly 15 million acres of forests and woodlands for all ownerships in the State. Again, however useful this information is now, it would be highly desirable to have this structural diversity mapped.

Use of the 4 C's to Determine Desired Conditions

Campbell and Bartos (2001) describe 4 C's used to determine desired conditions.

Commitment—devote the time and resources to allow the process to occur and mature.

Communication—talk and interact willingly and openly with each other.

Collaboration—promote intense and enthusiastic sharing of information.

Cooperation—work together; walk the talk; make it happen!

These concepts certainly apply to determining the use of IWFIA map data as well. The RMRS-Ogden IWFIA group continues to promote these 4 C's. During the past 3 years, Renee O'Brien (RMRS), Gretchen Moisen (RMRS), Tracey Frescino (RMRS), and Tom Edwards (Associate Professor, Utah State University) made three trips to the FNF Supervisor's Office in Richfield. A variety of different mapping projects were discussed and planned for the FNF. These efforts also led to Randy Schultz (USU graduate student) spending two summers on the Fishlake gathering wildlife habitat data for his project. Ogden-IWFIA researchers interacted on numerous occasions with more than a dozen Forest specialists. The FNF benefited from these associations and the resulting synergy with the IWFIA researchers. Forest specialists could not have completed various assessments without the IWFIA products.

Current Situation

Studies in the Beaver River Watershed

The FNF is fortunate to have several studies completed, ongoing, or proposed in the Beaver River watershed with about 123,000 acres administered by the Forest Service. Researchers from six units of RMRS and Utah State University have visited this watershed. The Beaver River drainage was the FNF's flagship watershed assessment for 2002. As a result of that assessment, nearly 20 projects were identified as ways to restore and sustain properly functioning conditions. The IWFIA group is working on structural diversity and wildlife habitat maps, as well as projects with ultra-low-level aerial photography and high resolution mapping from satellite images. The watershed is a focus area for a large fuels modeling project in addition to a tree-ring analysis of fire history. RMRS scientists also desire to study treatments in the pinyon/juniper type and forage reduction associated with decline of the aspen cover type.

Was it coincidence that all of these studies included portions of the Beaver River drainage? No, it did not just happen. The FNF began to focus attention on the drainage a few years ago. Since the FNF is considered by many to be a research-friendly study area, when RMRS scientists and university professors asked if there were any areas where a particular study might occur, the response from Richfield was usually the Beaver River watershed. And some of the studies are there because of serendipity. Whatever the reason, the result is that the FNF is amassing a substantial database about this biologically diverse as well as socially and economically important watershed.

Mapping of Spatially Explicit Information

We tie back to the elements of biodiversity and consider composition and structure again. Most forests have compositional diversity data mapped to some extent. However, many forests do not have forestwide maps more refined than for land type associations. The FNF and most other forests lack maps of structural diversity. However, an exception to this would be forests with predominantly timber resources.

The FNF is beginning the forest plan revision process. IWFIA products would be useful at many stages of the revision process. Such products will provide credible scientific underpinnings for the analyses that will lead to a revised land and resource management plan. Current needs for specific information include hazardous fuels treatments, fuel loadings, timber harvests linked to spruce beetle epidemics, and wildlife management indicator species (e.g., goshawk, cavity nesters, sagebrush guild, deer, and elk).

Resource Questions and Application Needs for IWFIA Products

Display Spatially Explicit Information for Structural Diversity

Vegetation cover is mapped for the entire Forest based on soils maps scaled at 1:24,000. The IWFIA data corroborated FNF Soil Scientist Mike Smith's forestwide existing vegetation/soils map done at a scale of 1:24,000 based on the documentation collected from various soil polygons. These vegetation/soils maps are used regularly for project evaluation and implementation. IWFIA researchers place high value on and have great interest in these soils maps. New IWFIA products might further corroborate these data layers and reinforce the concept of compositional diversity. Mapping of structural diversity is key! Specific examples are distribution of age classes by cover type for all vegetation types including the non-forest types and structure of shrub communities and woody understories. Possibly maps of historical fire patterns

and other disturbances could be derived which would allow function and process to be addressed.

Recently, IWFIA maps were used to stratify sample points for a new stand exam contract for the FNF. Forest specialists realize that not all IWFIA layers are equally useful. Some IWFIA layers appear linked in theme, concept, and display. Maps of biomass, volume, basal area, and density may be correlated and show essentially all the same. Some questions may require additional field investigation. For example, does the IWFIA volume layer equate to biomass or fuel loading as defined by fuels specialists?

Scientific Underpinnings for Forest Plan Revision

Forest planning and resource specialists anticipate that IWFIA maps and other products will provide scientific underpinnings for the forest plan revision process. For management indicator species (MIS), the measure is status and/or trends in populations, habitats, and ecological conditions (USDA Forest Service 2000b). "Selected species populations and habitats representing land and resource management plan objectives that will be tracked to measure progress toward the (2006) milestone" for the area that contains the FNF include aspen and sage grouse in the sagebrush-steppe habitats. Spatially explicit information would be valuable to help meet these MIS monitoring measures and milestones.

IWFIA maps that display information spatially will:

1. enhance our understanding of properly functioning condition and desired conditions
2. tie directly to aspects of fire, fuels, timber, and wildlife management
3. support the forest plan revision process

Summary

In addition to knowing *what* resources the FNF has, spatially explicit displays of *where* those resources occur will be beneficial for multiple issues. To the extent that such maps address structure, composition, and function, these products will link to discussions of biodiversity and sustainability within the framework of properly functioning condition. Such IWFIA products would relate to the resource questions and application needs that exist on the FNF. It will be important to continue to seek opportunities to promote synergy in the development and use of the new IWFIA spatial products that are becoming available.

Literature Cited

Bartos, Dale L.; Campbell, Robert B., Jr. 1998a. Decline of quaking aspen in the Interior West—examples from Utah. Rangelands. 20: 17–24.

Bartos, Dale L.; Campbell, Robert B., Jr. 1998b. Water depletion and other ecosystem values forfeited when conifer forests displace aspen communities. In: Potts, Donald F., ed. Proceedings of AWRA Specialty Conference, Rangeland Management and Water Resources; 27–29 May 1998; Reno, NV. Tech. Pub. Series TPS-98-1. Herndon, VA: American Water Resources Association: 427–434.

Campbell, Robert B., Jr.; Bartos, Dale L. 2001. Aspen ecosystems: objectives for sustaining biodiversity. In: Shepperd, Wayne D.; Binkley, Dan; Bartos, Dale L.; Stohlgren, Thomas J.; Eskew, Lane G., comps. Sustaining aspen in western landscapes: symposium proceedings; 13–15 June 2000; Grand Junction, CO. Proceedings RMRS-P-18. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 299–307.

Edwards, T.C., Jr.; Schultz, R.J., Jr.; Moisen, G.G.; Frescino, T.S. 2004. The application of FIA-based, data to wildlife habitat modeling: a comparative study. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Frescino, T.S.; Moisen, G.G. 2004. Predictive mapping of forest attributes on the Fishlake National Forest. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Jackson, Linda L.; Campbell, Robert B.; Chappell, Linda M.; Greenhalgh, Kevin P.; Shiverdecker, Wallace T.; Weaver, Douglas A. 1998. Fishlake National Forest prescribed natural fire plan. Richfield, UT: U.S. Department of Agriculture, Forest Service, Fishlake National Forest. 146 p.

O'Brien, Renee A. 1999. Comprehensive inventory of Utah's forest resources, 1993. Resour. Bull. RMRS-RB-1. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 105 p.

O'Brien, Renee A.; Brown, Susan B. 1998. Forest resources of the Dixie National Forest. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 9 p.

O'Brien, Renee A.; Collins, Dennis. 1997. Forest resources of the Uinta National Forest. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 9 p.

O'Brien, Renee A.; Pope, Reese. 1997. Forest resources of the Wasatch-Cache National Forest. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 9 p.

O'Brien, Renee A.; Tymcio, Ronald P. 1997. Forest resources of the Manti-La Sal National Forest. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 9 p.

O'Brien, Renee A.; Waters, Shirley H. 1998. Forest resources of the Fishlake National Forest. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 9 p.

O'Brien, Renee A.; Woudenberg, Sharon W. 1998. Forest resources of the Manti-La Sal National Forest. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 9 p.

Schultz, R.J., Jr.; Edwards, T.C., Jr.; Moiser, G.G.; Frescino, T.S. 2004. Development and validation of spatially explicit habitat models for cavity-nesting birds in Fishlake National Forest, Utah. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Terletzky, P.; Frescino, T.S. 2004. Integrating spatial components into FIA models of forest resources: some technical features. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

USDA Forest Service. 1998. Properly functioning condition rapid assessment process—October 23, 1998 version. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Region (www.fs.fed.us/r4/goshawk/r4-pfc-process.pdf). 33 p.

USDA Forest Service. 2000a. Properly functioning condition rapid assessment process—January 7, 2000 version. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Region (fsweb.r4.fs.fed.us/unit/vm/ecology/pfc_temp.html). 32 p.

USDA Forest Service. 2000b. USDA Forest Service strategic plan (2000 revision). Washington, DC: U.S. Department of Agriculture, Forest Service, Washington Office. 73 p.

Predictive Mapping of Forest Attributes on the Fishlake National Forest

Tracey S. Frescino and Gretchen G. Moisen[1]

Abstract.—Forest land managers increasingly need maps of forest characteristics to aid in planning and management. A set of 30-m resolution maps was prepared for the Fishlake National Forest by modeling FIA plot variables as nonparametric functions of ancillary digital data. The set includes maps of volume, biomass, growth, stand age, size, crown cover, and various aspen characteristics. Ancillary data layers included pre-classified TM data, raw TM bands, and topographic variables. Predictive models were built using automated multivariate adaptive regression splines (MARS), and refined using local knowledge and digital orthoquads (DOQs). Validation and application issues are discussed.

National forest planners must frequently make decisions using existing information (Campbell and O'Brien 2004). There is rarely time or resources to collect new data specific to each question encountered. Tabular summaries and analytical reports prepared by the Forest Inventory and Analysis program (FIA) have proven useful for past assessments, but there is an increasing need for spatially explicit delineations of forest data. For example, maps are needed to assess suitable wildlife habitat, marketable harvest areas, desired future conditions, and historical distributions of forest cover types.

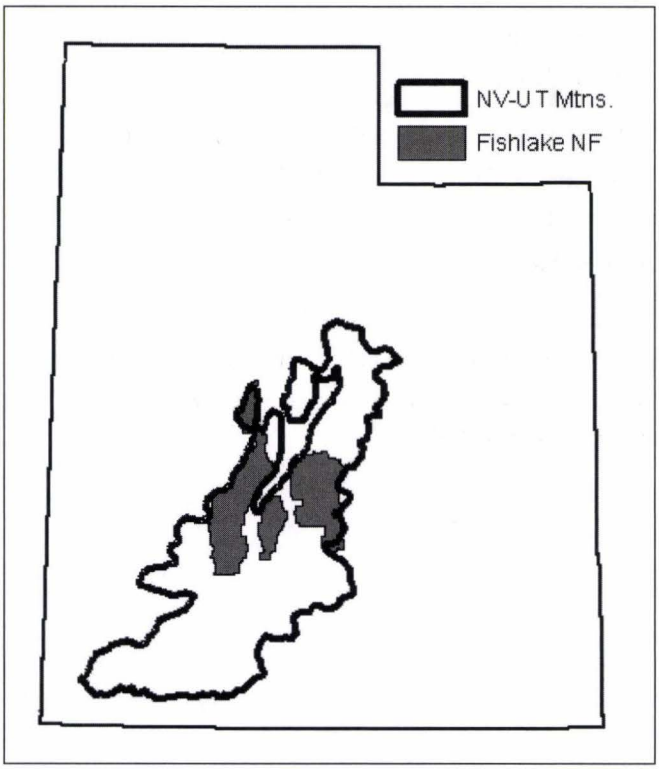
Our study demonstrates a method for generating spatially explicit maps of various forest attributes for use on national forests. The overall objective was to generate a series of maps to facilitate national forest management planning and to assist with a wildlife modeling study of cavity-nesting birds in aspen stands (Schultz 2002; Schultz *et al.* 2004; Edwards *et al.* 2002, 2004). Specifically, our objectives were to (1) build predictive models integrating FIA plot data with 30-m resolution digital data using multivariate adaptive regression splines (MARS) and geographical information systems (GIS) techniques; (2) refine and validate the models with statistical and visual error estimates; and (3) generate 30-m resolution maps of various FIA variables.

Methods

Study Area

The Fishlake National Forest comprises approximately 1,434,500 acres of land located in central Utah (fig. 1). It is a diverse forest with elevations ranging from less than 5,000 feet to over 12,000 feet. The forest supports a variety of vegetative cover types and forest resources. Pinyon-juniper cover types occur at low elevations and provide valuable habitat for deer, elk, and various small mammals and songbirds. Ponderosa pine and aspen cover types appear at higher elevations. Ponderosa pine provides valuable wildlife cover and is a valuable com-

Figure 1.—*Training data extents: Fishlake National Forest boundary and the Nevada-Utah Mountain ecoregion boundary.*



[1] Interior West Forest Inventory and Analysis, Rocky Mountain Research Station, U.S. Department of Agriculture, Forest Service, 507 25th Street, Ogden, UT 84401.

Table 1.—*FIA forest attributes, including units and description*

Forest Attribute (Alias)	Units	Description
Tree basal area (BALIVE)	Sq. ft./acre	Basal area of live trees 1 inch diameter and greater
Tree volume (NVOLTOT)	Cu. ft./acre	Net volume of live trees 5 inches diameter and greater
Tree biomass (BIOMASS)	Tons/acre	Woody biomass per acre of live trees 1 inch diameter and greater
Tree crown cover (CRCOV)	%	Crown cover of live trees 1 inch diameter and greater
Trees per acre (TPA)	# Trees	Trees per acre of live trees 1 inch diameter and greater
Stand age (STAGE)	Years	Weighted average age of the stand
Quadratic mean diameter (QMD)	Inches	Tree diameter based on the weighted average basal area of live timber trees 1 inch diameter and greater and live woodland trees 3 inches diameter and greater
Net annual growth (NGRWCF)	Cu. ft./acre	Annual net volume growth per acre of live growing-stock timber trees 5 inches diameter and greater and woodland trees 3 inches diameter and greater
Aspen presence (ASP)	Yes/no	Presence of aspen trees 1 inch diameter or greater
Aspen basal area (ASPBA)	Sq. ft./acre	Basal area of live aspen trees 1 inch diameter and greater
Percent aspen basal area	%	Percent basal area of live aspen trees 1 inch diameter and greater
Average tree height (TRHTAVG)	Feet	Average height of dominant or codominant trees
Snag density (SNAGNUM)	# Snags	Snags per acre of standing dead trees 5 inches diameter and greater
Aspen rot presence (ASPROT)	Yes/no	Presence of aspen disease

mercial tree species. Aspen is widely known to be prime wildlife habitat, affording beneficial cover, water, and food for a variety of wildlife species. Aspen cover types are being threatened by successional climax species, such as subalpine fir, white fir, and spruce, which are crowding out the aspen and diminishing the benefits to wildlife. Spruce-fir cover types occur at the higher elevations.

The Forest falls almost entirely within Bailey's (1980) Nevada-Utah Mountain ecoregion province, revised by Homer *et al.* (1997) (F-1). FIA data were available throughout this ecoregion. Because this ecoregion is ecologically similar to the Fishlake National Forest, we considered modeling forest characteristics for both areas.

Data

There were 836 forested locations within the Nevada-Utah Mountain ecoregion and 231 forested FIA locations within the Fishlake National Forest. We identified a set of eight FIA forest attributes to assist with management planning (tree basal area, tree volume, tree biomass, tree crown cover, trees per acre, quadratic mean diameter, stand age, and net annual growth), and a set of six additional variables needed for modeling aspen habitats for cavity-nesting birds (aspen presence, aspen basal area, percent aspen basal area, average tree height, snag density, and aspen rot presence). We used data collected on the FIA plots to compile individual tree measurements and combined them with stand variables to produce location-level summaries of all variables (table 1).

Data extraction and mining routines were performed within a GIS environment. We acquired a set of twelve 30-m resolution digital layers that would be appropriate for predicting forest attributes (table 2). Seven of these layers were based on 30-m resolution Enhanced Thematic Mapper (ETM) satellite data obtained through the Multi-Resolution Land Characteristics (MRLC) consortium. Three were raw spectral bands, one was a normalized difference vegetation index (NDVI) derived from the raw spectral bands, and the remaining three were classified ETM products generated by the Land Cover Characterization (LCC) program of the U.S. Geological Survey (USGS) Earth Resources Observation Systems (EROS) Data

Center (EDC) (Huang *et al.*, in press). The other five predictor variables were derived from 30-m resolution National Elevation Dataset (NED) digital elevation models (DEMs), including elevation, aspect, slope, hillshade, and topographic class. Elevation was extracted directly from the DEMs while aspect, slope, and hillshade were derived from the DEM using functions from the GRID module in ArcInfo GIS (ESRI Inc., Redlands, CA). The topographic class variable was derived from the DEM using a customized arc macro language AML (Zimmerman, unpublished data). The aspect variable was transformed from degrees to a symmetric radiation wetness index, calculated using the following formula (Roberts and Cooper 1989):

$$\text{Aspect} = \frac{1 - \cos(aspect - 30)}{2}$$

This transformation assigns the highest values to land oriented in a north-northeast direction, the coolest and wettest orientation in Utah. The hillshade variable was derived using an illumination angle of 225 degrees.

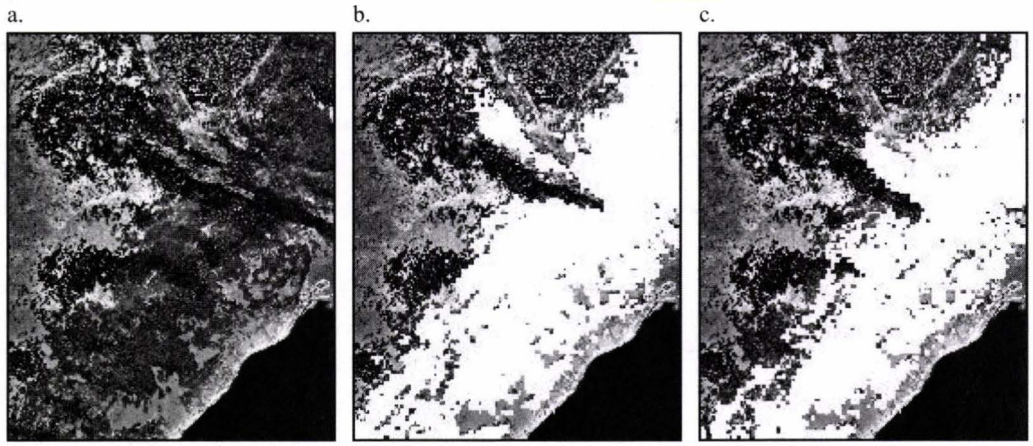
Models

Predictive models of various forest attributes were generated using Multivariate Adaptive Regression Splines (MARS) (Friedman 1991, Prasad and Iverson 2002, Steinberg *et al.* 1999). MARS is a flexible, nonparametric regression modeling tool that automatically finds the complex relationships between a response variable and a set of continuous and discrete predictors. MARS builds models by fitting numerous piecewise linear regressions, and approximates nonlinearity by allowing the slope of the regression lines to change over different intervals of the predictor space. These intervals are defined by *basis functions*, which are the building blocks of a MARS model. MARS starts by building a large and overly complex model with many basis functions. An optimal model is then found by deleting basis functions in order of least contribution to model performance. This prevents over-fitting and ensures that the mode will stand up to new data for prediction applications such as mapping. Features of MARS that make it particularly well suited to mapping forest attributes are that it handles both categorical and continuous variables, selects the relevant predictor

Table 2.—*Ancillary data predictor variables, including units and description*

Forest Attribute (Alias)	Units	Description
ETM Band 3 (ETMB3)	Brightness value (0-255)	Red (0.63 - 0.69 micrometers); June 2000–leaf on
ETM Band 4 (ETMB4)	Brightness value (0-255)	Near-infrared (0.76 - 0.90 micrometers); June 2000–leaf on
ETM Band 5 (ETMB5)	Brightness value (0-255)	Mid-infrared (1.55 – 1.75 micrometers); June 2000–leaf on
ETM NDVI (ETMVI)	0.0 – 1.0	Normalized Difference Vegetation Index; June 2000–leaf on
Classified ETM (LCC10)	10 classes	2:Nonforest; 10:Pinyon/juniper;15:Douglas-fir; 20:Ponderosa pine; 30:Spruce/fir; 35:Lodgepole; 50:Other western softwoods; 75:Aspen/birch; 85:Western oak; 90:Other western hardwoods (based on June 2000–leaf on ETM)
Classified ETM (LCC4)	4 classes	2:Nonforest; 41:Deciduous; 42:Evergreen; 43:Mixed (based on June 2000–leaf on ETM)
Classified ETM (LCC2)	2 classes	1:Forest; 2:Nonforest (based on June 2000–leaf on ETM)
Elevation (ELEV)	Meters	Elevation from mean sea level
Aspect (TRASP)	0 to 1	Transformed index representing radiation and wetness
Slope (SLP)	%	The rate of change from one cell to the next
Hillshade (HLSHD)	Brightness value (0-255)	Shaded relief considering shadows and an illumination angle of 225 degrees
Topoclass (TOPOCL)	4 classes	Classified to identify topographic features (1:Ridge;, 2:Slope; 3:Toe slope; 4:Valley bottom)

Figure 2.—*Visual assessment of aspen presence predictions compared to a DOQ. a. DOQ without predictions. b. Predictions based on a model built using the Nevada-Utah Mountain ecoregion data set; c. Predictions based on a model built using the Fishlake National Forest boundary data set. White represents the predicted aspen presence.*



variables and specifies their relationship with the response automatically, determines the level and nature of interactions as well as transformations, handles missing values, protects against over-fitting, and is fast and efficient for large data sets.

We examined the effect of using training data from the two different data extents, the Fishlake National Forest and the Nevada-Utah Mountain ecoregion. In all cases, we ran the models with a maximum of 100 basis functions, specified ten-fold cross-validation to select model degrees of freedom and prevent over-fitting, and allowed second-order interactions between predictor variables. Model performance was evaluated and refined by looking at R^2 and mean square error measures generated by MARS on a subset of the forest variables. We also visually assessed the model predictions using parallel screen displays of digital orthoquads (DOQs) and the output maps.

Maps

Maps were generated within a GIS environment. MARS output was converted to Arc Macro Language (AML) using Visual Basic (J. Nelson, unpublished data) and then run in Arc GRID. Thirty-meter resolution, spatially explicit maps were output for each FIA attribute. The nonforest class from the LCD classified-ETM product was used to mask the nonforest areas on the ground. Alternative approaches to applying MARS models to large geographic areas using Iterative Data Language are discussed in Terletzky and Frescino (2004).

Results

The models using the 836 training locations within the Nevada-Utah Mountain ecoregion performed better than the models using 231 training locations within the Fishlake National Forest in most cases. Table 3 shows the R^2 and MSE results from MARS for eight different FIA attributes, comparing the models built using different training data sets. The numbers in bold represent higher R^2 and lower MSE values, indicating better model fits. For five of the eight attributes, the R^2 values were higher when using the Nevada-Utah Mountain ecoregion data set. MSE values were lower when using the Nevada-Utah Mountain ecoregion data set for seven of the eight attributes.

Figure 2 shows an example of the visual assessment for models predicting aspen presence, comparing prediction results with what is displayed from a DOQ. The visual assessments of the predictions from the models built from the Nevada-Utah Mountain ecoregion data set appeared better than the predictions from the models built using the Fishlake National Forest data set.

Discussion and Conclusions

Predictive modeling is not an exact science. Many factors influence model performance. One is the extent of the training data set. With several examples and evaluation procedures, we

determined that the models performed "better" when using a larger data set. Although many of the data were outside the area of interest, the data were ecologically similar and significantly helped to establish functional relationships between the forest attributes and the ancillary data products.

What does "better" mean? Although we were able to objectively compare model performance using R^2 and MSE measures, further analysis is needed to validate the model predictions using independent data and true tenfold cross-validation procedures. More importantly, global measures of map accuracy often cannot capture what is obvious to a forest land manager wanting to use predictive maps in real-world applications. Further investigation is needed to build measures of utility into the picture.

As mentioned previously, one of the features of MARS is that it selects the relevant predictor variables. This allows us to see which predictor variables most influence the occurrence of the different forest attributes. Table 4 shows the predictor variables that were used to build the final models for eight forest attributes. The order of the variables corresponds to the relative importance of each in the model, or the amount of variance reduced by each. In general, the variables that seemed to have the most influence were the ETM raw spectral bands 5 and 3, elevation, the classified-ETM 10 and 4 classes, and the topographic class. This makes sense since band 5 (mid-infrared) characteristically indicates vegetation moisture and band 3 (red) responds to chlorophyll absorption. Elevation is a surrogate for temperature and moisture as well as the topographic class that distinguishes ridges from slopes from valley bottoms. The classified-ETM products would help distinguish differences between different forest classes, removing shadows and other features that the raw imagery may confuse.

Modeling forest attributes is an attempt to delineate characteristics in the landscape using available field data and ancillary resources, such as satellite imagery and topographic data. We assume there are significant relationships between these attributes and ancillary resources. Further research is needed to refine these relationships and obtain new ancillary products to build more accurate models.

Table 3.—*R^2 and MSE results from MARS for models built using the Fishlake National Forest (Fnf) data set and Nevada-Utah Mountain ecoregion (Uteco) data set (numbers in bold represent the best-fit model)*

Attribute	Training data set	R^2	MSE
BALIVE	Fnf	0.053	4,575.91
	Uteco	**0.284**	**3,012.26**
NVOLTOT	Fnf	0.419	1,255,907.10
	Uteco	**0.525**	**1,110,762.60**
BIOMASS	Fnf	0.348	372.43
	Uteco	**0.476**	**330.91**
CRCOV	Fnf	**0.438**	**290.15**
	Uteco	0.385	295.51
TPA	Fnf	0.335	110,635.50
	Uteco	**0.349**	**96,821.37**
Stage	Fnf	0.046	4,419.27
	Uteco	**0.114**	**3,335.20**
Trhtavg	Fnf	**0.739**	308.74
	Uteco	0.554	**195.20**
Aspba	Fnf	**0.478**	2,110.66
	Uteco	0.396	**1,956.01**

Table 4.—*Relevant variables contributing to model variance reduction*

Forest attribute	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7
BALIVE	ETMB5	ELEV	LCD10	LCD2	TOPOCL		
NVOLTOT	ELEV	ETMb3	ETMB5	LCD4	TOPOCL	LCD10	ETMB4
BIOMASS	ETMB5	ELEV	LCD10	TOPOCL			
CRCOV	ETMB3	LCD4					
TPA	LCD4	ELEV	ETMB5				
STAGE	LCD4	ELEV	ETMB5				
TRHTAVGN	LCD10	ETMB3	ELEV	TOPOCL			
ASPBA	LCD4	LCD10	ETMb3	ETMb5	ELEV	SLP	

National forest planners are enthusiastic about incorporating these spatially explicit products into their planning procedures and integrating them with other digital data to help understand the spatial diversity in the landscape and make decisions related to wildlife habitat, marketable harvest areas, desired future conditions, and so on. Wildlife modelers are also enthusiastic about adding spatially explicit maps of specific attributes into their models. These maps will provide valuable information about structural components of the forest and allow predictions of wildlife species, spatially across the landscape.

Literature Cited

Bailey, R.G. 1980. Description of the ecoregions of the United States. Misc. Publ. 1391. Washington, DC: U.S. Department of Agriculture.

Campbell, R.C., Jr.; O'Brien, R.A. 2004. Spatial information needs on the Fishlake National Forest: Can FIA help? In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Edwards, T.C., Jr.; Moisen, G.G.; Frescino, T.S.; Lawler, J.J. 2002. Modeling multiple ecological scales to link landscape theory to wildlife conservation. In: Bissonette, J.A.; Storch, I., eds. Landscape ecology and resource management: making the linkages. Covelo, CA: Island Press.

Edwards, T.C., Jr.; Moisen, G.G.; Frescino, T.S; Schultz, R.J. 2004. The application of FIA-based, data to wildlife habitat modeling: a comparative study. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Friedman, J. H. 1991. Multivariate adaptive regression splines. Annals of Statistics. 19: 1–141.

Homer, C.; Ramsey R.D.; Edwards, T.C., Jr.; Falconer, A. 1997. Landscape cover-type modeling using multi-scene TM mosaic. Photogrammetric Engineering and Remote Sensing. 63: 59–67.

Huang, C.; *et al.* 2002. Synergistic use of FIA plot data and Landsat 7 ETM+ images for large area forest mapping. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; Moser, John W., eds. Proceedings, 3rd annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MI: U.S. Department of Agriculture, Forest Service, North Central Research Station: 50–55.

Omernik, J.M. 1987. Map supplement: ecoregions of the conterminous United States. Annals of the Association of American Geographers. 77: 118–125 (map).

Prasad, A.M.; Iverson, L.R. 2002. Predictive vegetation mapping using a custom built model-chooser: comparison of regression tree analysis and multivariate adaptive regression splines. In: Proceedings of the 4th International conference on integrating GIS and environmental modeling: problems, prospects and research needs; 2000 September 2–8; Banff, Alberta, Canada.

Roberts, D.W.; Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: Ferguson, D.; Morgan, P.; Johnson, F.D., eds. Land classification based on vegetation: applications for resource management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station: 90–96.

Schultz, R.J., Jr. 2002. Scale dependency in habitat selection by cavity nesting birds in Fishlake National Forest. Logan, UT: Utah State University. 100 p. M.S. thesis.

Schultz, R.J., Jr.; Edwards, T.C.; Moisen, G.G.; Frescino, T.S. 2004. Development and validation of spatially explicit habitat models for cavity-nesting birds in Fishlake National Forest, Utah. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Steinberg, D.; Colla, P.L.; Martin, K. 1999. MARS users guide. San Diego, CA: Salford Systems.

Terletzky, P.; Frescino, T.S. 2004. Integrating spatial components into FIA models of forest resources: some technical aspects. In: McRoberts, R.E.; *et al.,* eds. Proceedings, 4th annual forest inventory analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

TIGER 2000 and FIA

Joseph McCollum and Dennis Jacobs[1]

Abstract.—The legal foundations of the FIA (Forest Inventory and Analysis) program are laid out. Upon those foundations are built a geographical definition of the United States and its components, and how applying that definition might change from decade to decade. Along the way, the American system of weights and measures as well as the unusual geography of the Commonwealth of Virginia are explained. Some recommendations are offered for the FIA program.

This paper is primarily about the geography of the United States as it pertains to the FIA program. Some of the details discussed may seem unimportant but are included in the interest of thoroughness.

Several maps (U.S. Census Bureau 2000a) appear throughout the paper. To conserve space, figure 1 serves as a common key.

Legal Foundations

The Agricultural Research, Extension, and Education Reform Act of 1998, Public Law 105-185, was integrated into the United States Code as 16 U.S.C. 1642(e). The expanded citation is Title 16, Chapter 36, Subchapter II, Section 1642, Subsection (e). That law mandates the FIA program to survey the forest resources of the United States. Later in the same subchapter, 16 U.S.C. 1645(f) (Office of the Law Revision Counsel 2000) defines what the United States is:

> *For the purposes of this subchapter, the terms "United States" and "State" shall include each of the several States, the District of Columbia, the Commonwealth of Puerto Rico, the Virgin Islands of the United States, the Commonwealth of the Northern Mariana Islands, the Trust Territory of the Pacific Islands, and the territories and possessions of the United States.*

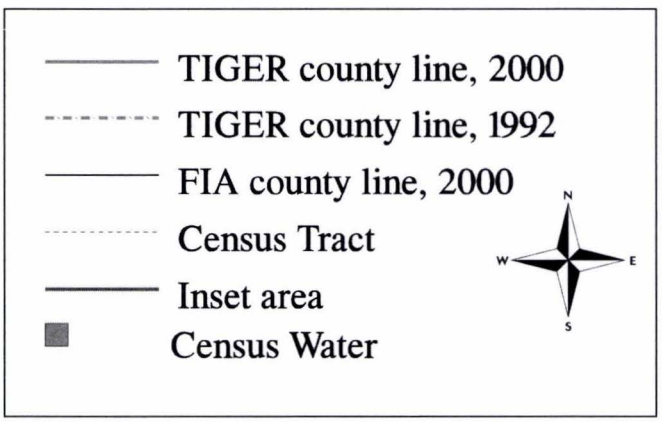
This provision was part of the Forest and Rangeland Renewable Resources Research Act of 1978 (Public Law 95-307), which became law on June 30, 1978.

The TTPI (Trust Territory of the Pacific Islands), entrusted to the United States by the United Nations in 1947, no longer exists. It consisted of what is now the NMI (Northern Mariana Islands), the RMI (Republic of the Marshall Islands), the FSM (Federated States of Micronesia), and Palau (The Republic of Palau). The NMI became a commonwealth of the United States in January 1978. The FIA program has announced its intention to survey the TTPI as it was defined at the time of the passage of Public Law 95-307 (USDA 1999).

The territories and possessions referenced in 16 U.S.C. 1645(f) are listed in table 1 (Bureau of the Census 1994, Central Intelligence Agency 2002). They are all in the Pacific Ocean except for Navassa Island in the Caribbean Sea. Also listed are the land and water area of each in mi^2 (square miles) and km^2 (square kilometers). Surface water estimates for the smaller possessions are not available, but maps and tables (Central Intelligence Agency 2002) show no inland water.

The Census Bureau (Bureau of the Census 1971) published the area of the Trust Territory at 717 mi^2 of land and 7,772 mi^2 of water. Data from Bryan (1971) suggest this estimate was primarily coastal water and did not include territorial sea. After the NMI became a commonwealth, land estimates were revised to 533 mi^2 for the TTPI and 184 mi^2 for the NMI

Figure 1.—*Key for maps.*



[1] Computer Specialist and Research Forester, U.S. Department of Agriculture, Forest Service, Southern Research Station, 4700 Old Kingston Pike, Knoxville, TN 37919 and 201 Lincoln Green, Starkville, MS 39759, respectively.

(Bureau of the Census 1981). They did not issue surface water estimates, but maps (Bureau of the Census 1982) indicate little inland water, if any. In 1986, the RMI and the FSM gained independence. In the next decade, the Census Bureau's estimate for Palau was 177.3 mi^2 of land and 452.6 mi^2 of water (Bureau of the Census 1994), including 40.1 mi^2 of inland water. The estimates for the NMI were revised to 179.0 mi^2 of land and 1,770.9 mi^2 of water, including 2.2 mi^2 of inland water. In 1994, Palau gained independence.

The latest estimates from the CIA (Central Intelligence Agency) report the RMI at 181.3 km^2 (70.0 mi^2) of land, the FSM at 702 km^2 (271 mi^2), and Palau at 458 km^2 (177 mi^2) of land (CIA 2002). The same source reports no inland water.

TIGER

TIGER files (Topologically Integrated Geographic Encoding and Referencing files) are the long-awaited precise digitized boundaries of census land and census water, first released in 1992. Upon that version of TIGER was laid the hexagon grid used for placing FIA plots. A hex center was assigned to a state if it landed in that state according to TIGER 1992.

The data are hierarchical, with States and State-equivalents at the top of the hierarchy. County-equivalents are one level below. In the several States, county-equivalents are counties, but also independent cities in Maryland, Missouri, Nevada, and Virginia, as well as parishes in Louisiana, and boroughs, census divisions, and the Municipality of Anchorage in Alaska. The District of Columbia is not divided into county-equivalents. In the territories, county-equivalents are municipios in Puerto Rico, islands in the Virgin Islands, municipalities in the NMI, and islands and districts in American Samoa. The *Census of Agriculture* (NASS 1996) reports Guam by election districts, although TIGER views such districts as minor civil divisions and all of Guam as one county-equivalent. To date, the only possession smaller than American Samoa for which TIGER files are produced is the Midway Islands, in one county-equivalent.

Several levels below the county-equivalent level are census tracts and beneath that level are census block groups and census blocks. Further details may be found in the Census 2000 TIGER/Line Technical Documentation (U.S. Census Bureau 2000b).

According to the *Geographic Areas Reference Manual* (U.S. Census Bureau 1994), area estimates were calculated from TIGER, but no further details were given, such as projection information. Raw TIGER data are in latitude and longitude. To calculate acres, one must project the data, at least indirectly.

There are many different projections. Since the users are interested in total surface area, it makes sense to use an equal-area projection. The Cylindrical Equal Area projection preserves cardinal directions in its equatorial aspect.
The Lambert Azimuthal Equal Area projection, based on a plane, preserves area and distance from the projection origin. Far away from the origin, the projection starts to bend back on itself and thus should not be used for an area much larger than a continent. The Albers Equal Area projection is based on a cone. It appears that for many counties in the conterminous United States, the Albers Equal Area projection, North American Datum of 1983, GRS (Geodetic Reference System) 1980 Spheroid, standard parallels of 45° 30' N. and 29° 30' N., with projection origin at 96° W. and 23° N., gives results that are nearly equal to those in the gazetteer.

Other choices are available. Although it is not quite equal area, State Plane is popular in the land surveying community. One caveat is that in South Carolina, "State Plane Feet" means International Feet, although acres are still in U.S. Survey units. As with UTM (Universal Transverse Mercator), the zones do not stitch together.

Although the exact amount of surface area varies somewhat with the choice of projection, TIGER files confirm 2.5 mi^2 (6.4 km^2) of land, no inland water, and estimate about 140 mi^2 (360 km^2) of coastal water and territorial sea for the Midway

Figure 2.—*Suggested Metric Equivalents table.*

Metric Equivalents (**bold** figures are exact)
1 *foot* = 1200/3937 meter = 0.304 800 610 meter
1 *acre* = 4,046.87 square meters or 0.404 687 hectare
1 inch = 2.54 centimeters or 0.0254 meter
1 foot = 0.3048 meter
breast height = 1.37 meters above ground
1 square foot = 929.03 square centimeters or 0.0929 square meter
1 cubic foot = 0.028 317 cubic meter
1 square foot basal area per *acre* = 0.229 567 square meter per hectare
1 pound = 0.453 592 37 kilogram
1 ton = 0.907 184 74 metric ton
U.S. Survey units are in *italics*. International units are in roman.

Islands. The other small possessions of the United States are about as compact as the Midway Islands; for them, one would expect a similar amount of coastal water and territorial sea.

One might think that the various equal-area projections might give identical answers, at least for identical spheroids. They would for sufficiently densified arcs, but GIS (Geographic Information Systems) programs assume a straight line between the points of a polygon in the current projection.

Area Measurement

Since one can get different answers depending on choice of projection as well as choice of software and computer platform, it might be best to see what areas the Census Bureau has published.

The *Area Measurement Reports* (Bureau of the Census 1970) listed land and inland water for each county-equivalent to a 10th of a square mile, as well as the number of square miles of offshore water (coastal water, territorial sea, and Great Lakes water) for each state.

The *Census of Agriculture* (e.g., NASS 1999a) publishes estimates of land in a county-equivalent to the nearest U.S. survey acre for most county-equivalents. Many independent cities in Virginia are not listed.

The *Population and Housing Unit Counts*, *CPH-2* (Bureau of the Census 1990) included land estimates to the nearest 10th of a square mile and 10th of a square kilometer for each county-equivalent in the 50 States, the District of Columbia, Puerto Rico, as well as the total for other territories and possessions. The data were published electronically at the census tract level (*CPH-3*) wherein land estimates were published to the nearest 100th of a square mile for each census tract.

However, the Census Bureau (U.S. Census Bureau 2001a, b) now publishes estimates of land and water to the nearest square meter in Summary Files 1 and 2. In the previous decade, such estimates were published to the nearest thousand square meters. A simplified version of this database, with the areas of county-equivalents reported in square meters and square International Miles, may be found on the Census Bureau's Web site, at www.census.gov/geo/www/gazetteer/places2k.html.

During the 1990s, FIA used an internal database. It closely followed but did not necessarily match the gazetteer or the *Census of Agriculture*.

The most precise units for area are in the gazetteer. It is in square meters, and converting it to acres is more difficult than it first appears to be.

Years before there was a National Biological Survey, there was the National Bureau of Standards. In 1988, it became the National Institute of Standards and Technology, or NIST. The agency is responsible for governing the weights and measures in the United States. In 1959, it offered refined values for the yard and the pound (National Bureau of Standards 1959). It defined the foot to be 0.3048 of a meter. Previously, it had been defined as 1200/3937 of a meter. The new unit was named the "International Foot" and the old unit would be called the "U.S. Survey Foot." Similarly, 1 International Yard was 3 International Feet, and 1 U.S. Survey Yard was 3 U.S. Survey Feet. Most measurements were to be made with International units, but geodetic measurements were to be made in Survey units. The 1959 memo envisioned retirement of the U.S. Survey units, but even now, the *NIST Handbook 44* recognizes an acre in the U.S. Survey system but not in the International (Butcher *et al.* 2001).

The acre based on the International Foot is exactly 4,046.856 4224 m^2 but is often reported to two or three significant digits beyond the decimal point. It is recognized by several standards boards around the world, including the Land

Table 1.—*Territories and possessions of the United States*

Territory	Land area		Water area	
	km^2	mi^2	km^2	mi^2
Guam	543.5	209.8	934.4	360.8
American Samoa	200.3	77.3	1310.4	505.9
Palmyra Atoll	11.9	4.6	(NA)	(NA)
Wake Island	6.5	2.5	(NA)	(NA)
Midway Islands	6.4	2.5	(NA)	(NA)
Navassa Island	5.2	2.0	(NA)	(NA)
Jarvis Island	4.5	—	(NA)	(NA)
Johnston Atoll	2.8	1.1	(NA)	(NA)
Howland Island	1.6	—	(NA)	(NA)
Baker Island	1.4	—	(NA)	(NA)
Kingman Reef	1.0	0.4	(NA)	(NA)

(NA) not available

Information New Zealand (www.linz.govt.nz/staticpages/dcdb/dataquality.htm), Measurements Canada, and the United Kingdom's National Weights and Measures Laboratory (www.nwml.gov.uk/consumer/units.asp). However, this acre is sometimes cited in American sources such as the CRC Press (Beyer 1978), the Army Corps of Engineers (Perrier *et al.* 1980), and even FIA (e.g., Smith *et al.* 2002). On the other hand, such sources as *The World Almanac* (Famighetti 1997) explain the difference between International and U.S. Survey units.

Breast height is 4.5 International Feet, exactly 1.3716 meters. While this number does round to 1.4 meters, breast height should be reported to at least three significant digits to reflect the precision with which the measurement is taken.

Figure 2 shows a Metric Equivalents table that incorporates the preceding recommendations, with decimals formatted according to the guidelines of the *U.S. Government Printing Office Style Manual* (U.S. Government Printing Office 2000).

Virginia

Although it refers to itself as the "Commonwealth of Virginia," Virginia is one of the several States and not a Commonwealth in the sense that Puerto Rico and the NMI are.

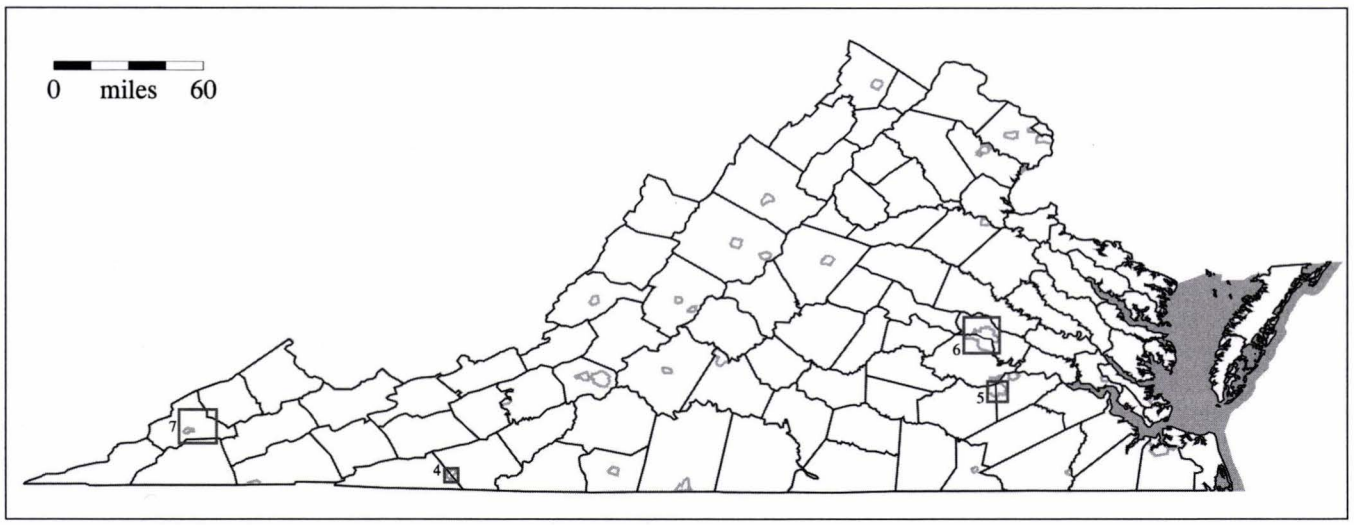
FIA prefers to report county acreages along traditional county-equivalent lines rather than legal ones, as shown in figure 3. This convention is not purely an invention of the FIA program; a similar map may be found on the Official Commonwealth of Virginia Home Page site, at www.vipnet.org/portal/images/vamap.jpg. Some independent cities are retained; most others are dissolved into surrounding counties or independent cities. There are three exceptions: the cities of Galax, Richmond, and Petersburg. These cities are each split between two legal counties. The distribution of land and water in square meters is shown in table 2.

One interesting result in this table is that GIS shows a disproportionately higher amount of census water in the western part of Petersburg than in the eastern part. It also shows the proportion of census water in the northern part of Richmond approximately equal to that in the southern part. The GIS technique is better than those used in the 1992 survey, where nomi-

Table 2.—*Split independent cities*

	Census Land m^2	Census Water m^2
Carroll	12,020,499	—
Grayson	9,290,641	—
Galax	21,311,140	—
Dinwiddie	24,176,386	565,298
Prince George	35,084,580	224,164
Petersburg	59,260,966	789,462
Chesterfield	87,962,551	3,692,453
Henrico	67,618,812	2,725,776
Richmond	155,581,363	6,418,229

Figure 3.—*Virginia.*



nal proportions were calculated from unrectified maps and photos and also nominal proportions were applied to census land and census water without regard to which part of the independent city they were in. Direct calculation in GIS clearly gives more accurate results.

Detailed maps of Galax, Petersburg, and Richmond are shown in figures 4, 5, and 6, respectively. Census blocks were assigned to FIA counties based on which side of the historical line the centroid was. In the case of Galax, the historical line is the line between Carroll County and Grayson County extended linearly. Census blocks whose centroids were west of this line are to be tabulated with Grayson County while those east of the line are to be tabulated with Carroll County. In the case of Petersburg, the historical line is more obscure, but it runs approximately north from the point of intersection between Petersburg, Dinwiddie County, and Prince George County. Census blocks whose centroids fall east of this line are tabulated with Prince George County while those whose centroids fall west of the line are tabulated with Dinwiddie County. In the case of Richmond, the historical line was maintained in modern-day TIGER, along census tract rather than just census block lines. The line is the James River, which is not a line, but a double-line stream. Census tracts south of this line are tabulated with Chesterfield County, while those north of the line are tabulated with Henrico County.

Another point is that Census Water polygons can be smaller than 4.5 acres, a fact that conflicts with the FIA national core field guide (USDA 2002). Among the original sources of TIGER data were 1:100,000 Digital Line Graphs of the USGS (United States Geological Survey). The standards for those data did allude to 4.5 acres as a minimum size for a water polygon, but they also said, "In arid and semiarid areas, the presence and location of water is important as a means of orientation. In these areas, as many hydrographic features as possible should be shown" (USGS 1991). While that source supports 200 feet as the minimum width of a double-line stream, the Census Bureau used other sources of data (most notably their own Metropolitan Map Series) in constructing TIGER files. These data did not necessarily adhere to the lower bound for water polygons of 4.5 acres nor to the lower bound of double-line streams of 200 feet.

TIGER 1992 vs. TIGER 2000

Apart from Virginia, three other States (Alaska, Maryland, and Montana) had jurisdictional changes at the county-equivalent level during the 1990s. However, there were many changes to the TIGER database between 1992 and 2000. These changes appear to have been digitizing errors being corrected. Luckily, no hex centers in the Southern Station switched States. A few hex centers went from territorial sea to international water. Figure 7 indicates two possible sites in Wise County, Virginia, that changed counties during the 1990s due to changes in the

Figure 4.—*Galax.*

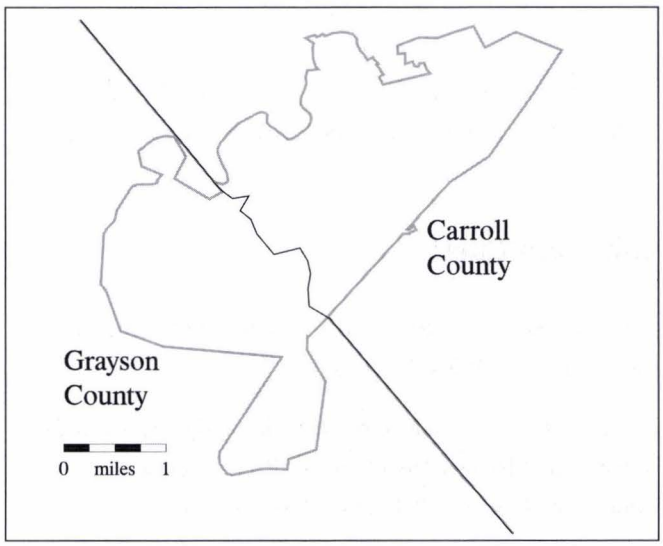


Figure 5.—*Petersburg.*

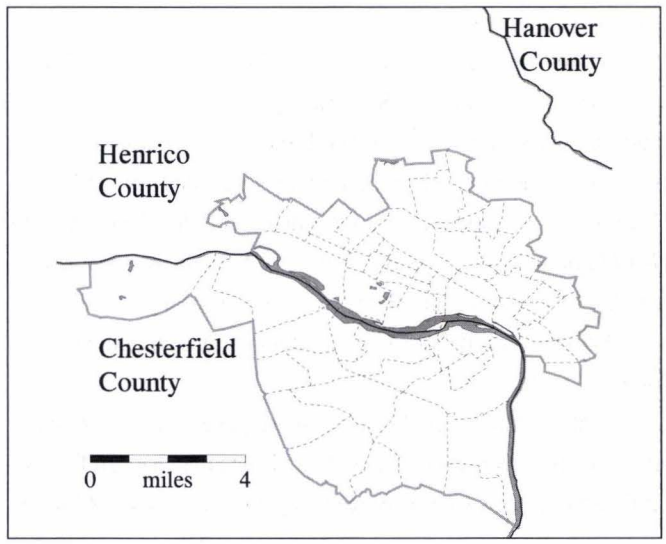


Figure 6.—*Richmond.*

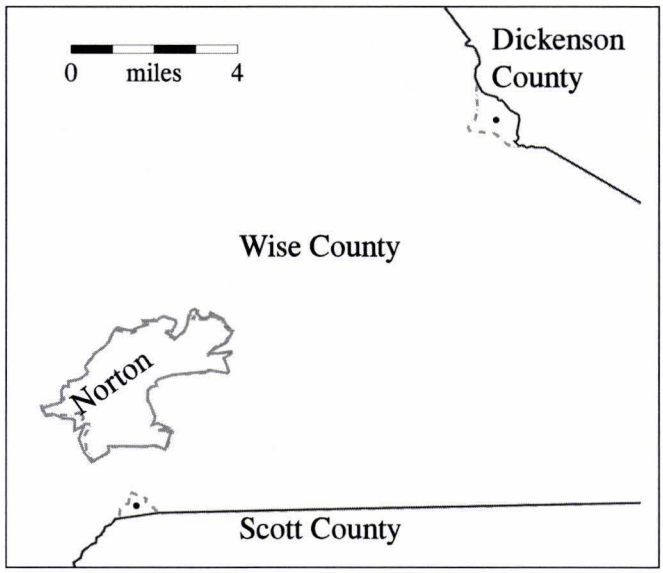
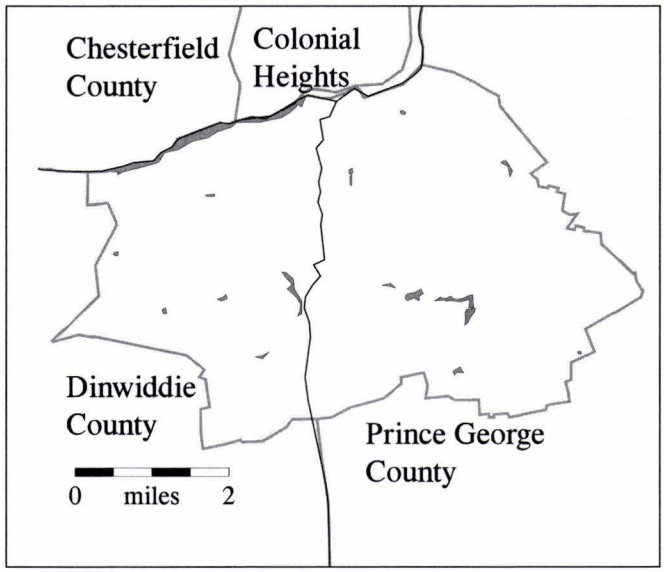


Figure 7.—*TIGER 1992 vs. TIGER 2000.*



TIGER database. Nationwide, there are hundreds, if not thousands, of such sliver polygons. Tracking them would be a monumental task.

Summary and Recommendations

First, logistical difficulties may make areas such as Navassa inaccessible. In such areas, photointerpretation should be done by remote sensing. A site visit to Navassa could be reconsidered if its spectral signatures were significantly different from accessible areas of Puerto Rico and the U.S. Virgin Islands.

Second, a Metric Equivalents table such as in figure 1 should be adopted.

Third, the program should adopt the most precise estimates of Census Area available. Specifically, those are the gazetteer and Summary Files for the several States, the District of Columbia, and Puerto Rico; Summary Files only for the U.S. Virgin Islands, American Samoa, Guam, and the NMI. The *CIA World Factbook* is a possible source for current area estimates of other possessions and freely associated nations.

Fourth, Virginia's FIA county delineations should be as outlined in this paper.

Fifth, a policy should be constructed governing what happens if an FIA plot switches counties. This change may happen for several reasons: 1. the Census Bureau may correct the TIGER line, 2. the GIS analyst may have generated a plot coordinate based on a faulty algorithm, 3. the GIS analyst may have relied on coordinates that were inaccurate, 4. the GIS technician may have used a faulty base map or lost his or her place while digitizing, 5. coordinates from GPS (Global Positioning System) units may have been collected or transcribed incorrectly, 6. finally, the field crew may notice that the plot is in the incorrect county.

Acknowledgments

The authors thank the Virginia Department of Forestry for their review of the geographical methods for their State, as well as Dr. Thomas Brandeis and Dr. Eileen Helmer of the International Institute of Tropical Forestry for their review of geographical methods in the tropics, and especially Elizabeth LaPoint of the FIA National GIS Service Center for her review.

Literature Cited

Beyer, W. ed. 1978. CRC standard mathematical tables. Boca Raton, FL: CRC Press, Inc. 613 p.

Bryan, E.H. 1971. Guide to place names in the Trust Territory of the Pacific Islands. Honolulu, HI: Pacific Science Information Center, B.P. Bishop Museum. 406 p.

Bureau of the Census. 1970. Area measurement reports. Washington, DC: U.S. Government Printing Office.

Bureau of the Census. 1971. Statistical abstract of the United States. Washington, DC.

Bureau of the Census. 1981. Statistical abstract of the United States. Washington, DC.

Bureau of the Census. 1982. 1980 census of population. Volume 1. Characteristics of the population. Chapter A. Number of inhabitants. Part 57B. Trust Territory of the Pacific Islands excluding the Northern Mariana Islands. Washington, DC.

Bureau of the Census. 1990. 1990 CPH-2-1. 1990 census of population and housing, population and housing unit counts, United States. Washington, DC: U.S. Government Printing Office. 834 p.

Bureau of the Census. 1994. Geographic areas reference manual. Available online at http://www.census.gov/geo/www/garm.html.

Butcher, T.; Grimes, T.; Williams, J., eds. 2001. Specifications, tolerances, and other technical requirements for weighing and measuring devices, NIST handbook 44 (various editors over the years). Available online at: http://www.nist.gov/public_affairs/pubs.htm.

Central Intelligence Agency. 2002. CIA world factbook, 2002. Washington, DC: U.S. Government Printing Office.

Famighetti, R., ed. 1997. The world almanac and book of facts 1997. Mahwah, NJ: World Almanac Books. 975 p.

National Agricultural Statistics Service. 1999a. 1997 census of agriculture. Virginia State and County Data. Volume 1, geographic area series part 46. 658 p.

National Agricultural Statistics Service. 1999b. 1998 census of agriculture. Guam. Volume 1, geographic area series part 53. 70 p.

National Bureau of Standards. 1959. Refinement of values for the yard and the pound. F. R. Document 59-5442. 2 p.

Office of the Law Revision Counsel. 2000. United States Code, 2000 edition, containing the general and permanent laws of the United States. Washington, DC: U.S. Government Printing Office.

Perrier, E.; Llopis, J.; Spaine, P. 1980. Area strip mine reclamation using dredged material: a field demonstration. Vicksburg, MS: U.S. Army Engineer Waterways Experiment Station. 6 p.

Smith, W. Brad; Vissage, John S.; Darr, David R.; Sheffield, Raymond M. 2002. Forest resources of the United States, 1997, METRIC UNITS. Gen. Tech. Rep. NC-222. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 127 p.

U.S. Census Bureau. 2000a. Census 2000 TIGER/Line files. [machine-readable data files]. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2000b. Census 2000 TIGER/Line technical documentation. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2001a. Census 2000 summary file 1, United States. Washington, DC: U.S. Census Bureau. Available online at: http://www.census.gov/PressRelease/www/2001/sumfile1.html.

U.S. Census Bureau. 2001b. Census 2000 summary file 2, United States, Washington, DC: U.S. Census Bureau. Available online at: http://www.census.gov/PressRelease/www/2001/sumfile2.html.

U.S. Department of Agriculture. 1999. A strategic plan for forest inventory and monitoring. Washington, DC: U.S. Department of Agriculture, Forest Service.

U.S. Department of Agriculture. 2002. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.6. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis.

U.S. Geological Survey. 1991. Standards for 1:100,000-scale quadrangle maps. Part 3: feature specifications and compilation. 132 p.

U.S. Government Printing Office. 2000. United States Government Printing Office style manual. Washington, DC: U.S. Government Printing Office. 326 p.

Regression and Geostatistical Techniques: Considerations and Observations from Experiences in NE-FIA

Rachel Riemann[1] and Andrew Lister[2]

Abstract.—Maps of forest variables improve our understanding of the forest resource by allowing us to view and analyze it spatially. The USDA Forest Service's Northeastern Forest Inventory and Analysis unit (NE-FIA) has used geostatistical techniques, particularly stochastic simulation, to produce maps and spatial data sets of FIA variables. That work underscores the importance of generating uncertainty information along with the modeled estimates, the value of incorporating additional satellite and other data into the modeling, and the need to understand the characteristics of the output data set. In our study, we investigated three questions: Does spatial structure matter when satellite-derived and ancillary spatial data sets are incorporated into the modeling of forest attributes? If we use a modeling technique such as multiple linear regression, how do we calculate or estimate the uncertainty? And what are the characteristics of the output data set with respect to the original sample data and the ancillary data used?

Background

Spatial depictions of forest variables improve our understanding of the forest resource by allowing us to view and analyze it spatially and ask questions such as: How are things distributed spatially, and how are they related to other social, environmental, and historical patterns? Estimates in small areas may be improved because additional, relevant information is incorporated into the modeling and estimation. Two extremely valuable data sources for mapping forest attributes are the Forest Inventory and Analysis (FIA) plot data and satellite-derived imagery. FIA data contain an enormous amount of information on a large number of sample plots collected in an unbiased manner and spread relatively uniformly across nearly the entire United States. Many ecosystems, small and large, and much of the variation within them are captured. In addition, satellite sensors capture data at every location (at various resolutions), and the resulting imagery often is strongly related to many of the forest variables we are interested in. For example, in a New Jersey study, we found that the correlation between basal area and several imagery-based satellite layers is high (table 1). In this paper we review the lessons learned from using the geostatistical technique sequential Gaussian conditional simulation (SGCS) to model the relative basal area of individual tree species, and investigate the use of multiple linear regression to model similar variables using satellite-based data sets and other available spatial layers.

Geostatistical Techniques

Much of the spatial modeling work in NE-FIA has been conducted with plot data and geostatistical techniques. For exam-

Table 1.—*Pearson's Correlation Coefficients between three tassel cap bands for each of three seasons (2000 data) and normal–score transformed relative basal-area values for each of three species variables being modeled (1996 data; all values are significant ($p< 0.05$), and bold values are more highly significant ($p<0.005$)*

Item	Pitch pine	Red maple	Total coniferous basal area
Spring_bright	0.26	-0.17	**-0.43**
Spring_green	**0.49**	**-0.23**	**0.61**
Spring_wet	**0.39**	—	**0.60**
Summer_bright	**-0.57**	**0.30**	**-0.67**
Summer_green	**-0.63**	**0.34**	**-0.58**
Fall_bright	**-0.56**	0.19	**-0.71**
Fall_green	**-0.54**	0.21	**-0.53**

[1] Research Forester/Geographer, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, c/o USGS, 425 Jordan Road, Troy, NY 12180.
[2] Forester, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, 11 Campus Blvd., Newtown Square, PA 19073.

Figure 1.—*Like kriging, sequential Gaussian conditional simulation (SGCS) uses a model of the spatial structure present to estimate values at unknown locations: a) original plots (darker red are increasing values of %ba of hemlock); b) correlogram calculated from the sample plot data; c) model used to describe the structure depicted in the correlogram; d) a resulting output map of modeled values (darker green are increasing values of %ba of hemlock).*

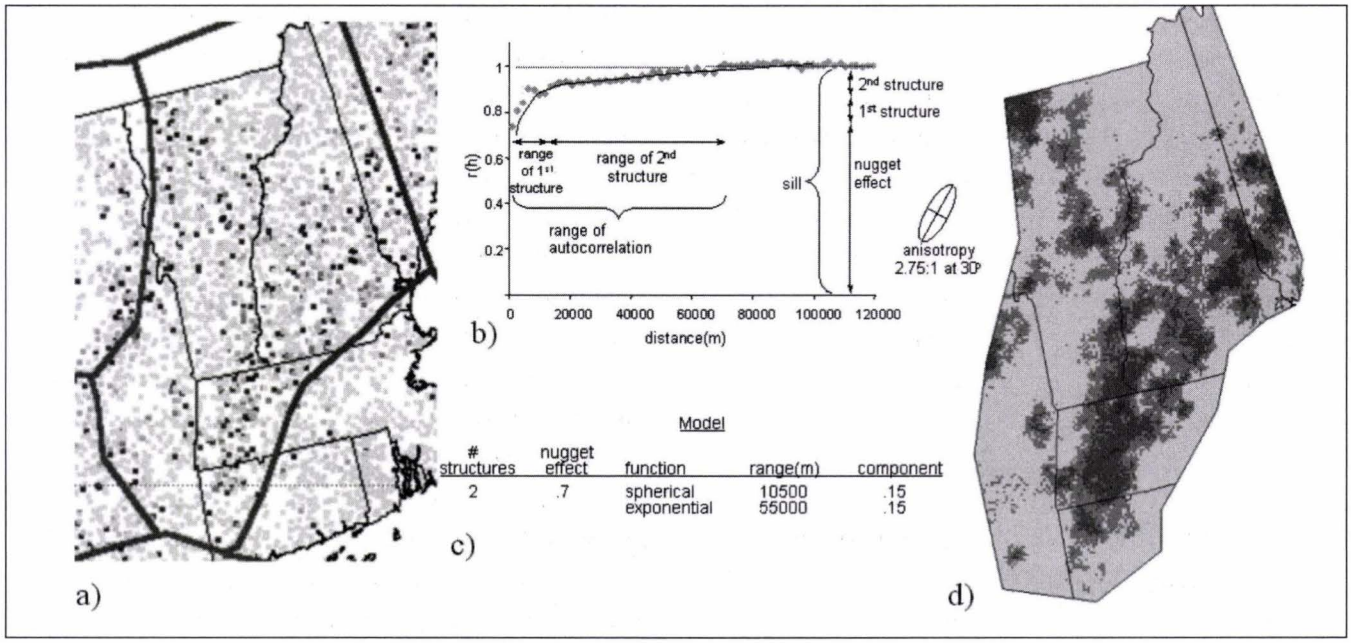
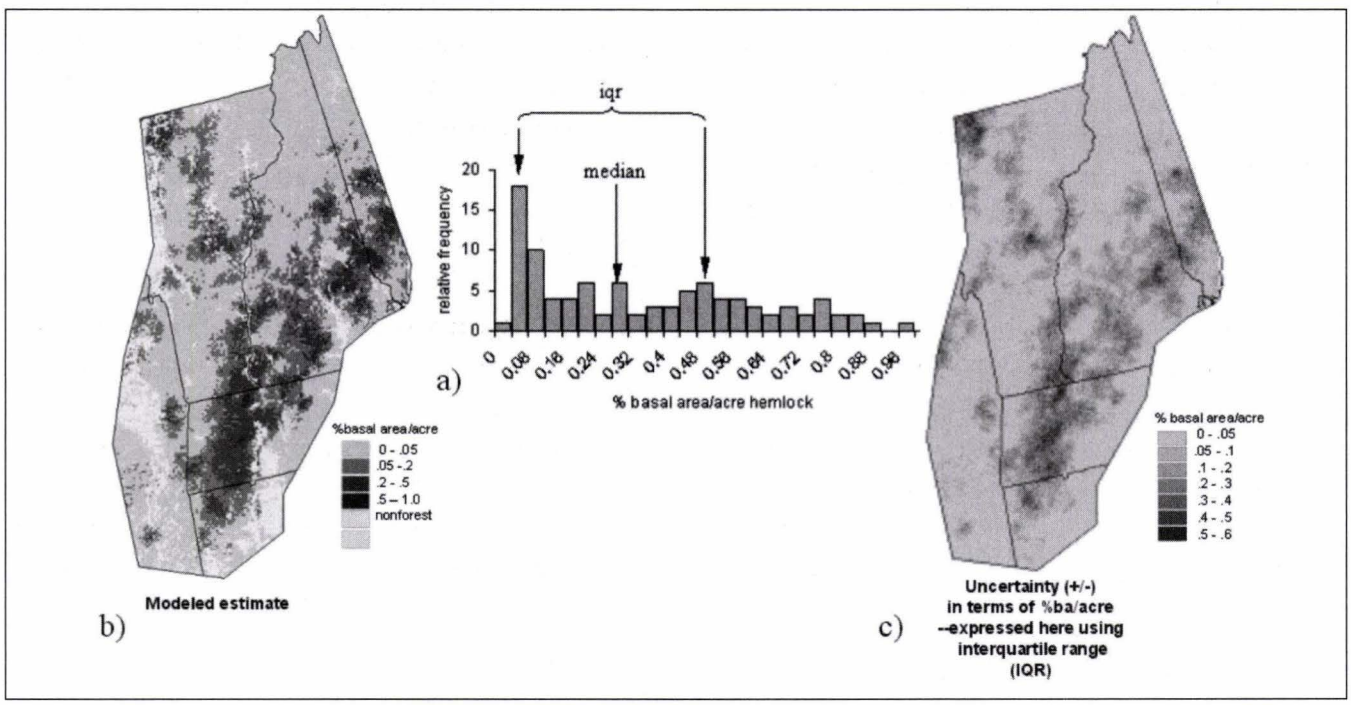


Figure 2.—*SGCS creates a distribution of possible values for each pixel (one generated with each simulation) from which the user can easily extract a clear measure of the uncertainty of each local estimate. a) the distribution of values at a single, randomly chosen cell; b) the modeled estimate where the value at the 65% percentile was chosen for each pixel (with nonforest areas masked out); and c) the value of the interquartile range (iqr) at each pixel, representing the range of uncertainty associated with each modeled value.*



ple, species distribution maps were created using SGCS (Riemann *et al.* 1997, Riemann and Lister,[3] http://www.fs.fed.us/ne/fia/spatial/specdist/species_dist.html). Like kriging, SGCS uses a model of the spatial structure present to estimate values at unknown locations (fig. 1). However SGCS differs in that it is a stochastic simulation technique that uses a random function to incorporate uncertainty into the model (Rossi *et al.* 1993). This creates a distribution of possible values for each pixel (one generated with each simulation) from which the user can easily extract a clear measure of the uncertainty of each local estimate (fig. 2). In addition, depending on the summary statistic chosen as the estimate (e.g., mean, median, or another percentile) this technique provides spatial output with a more realistic depiction of heterogeneity and is more effective than kriging in retaining original data characteristics.

Advantages and Limitations of SGCS

The use of SGCS offers several advantages. First is the way Monte Carlo techniques offer a clear depiction of the model uncertainty, which reflects both sample intensity and variation in the available data. This feature is valuable because some uncertainty is inevitable in all modeled output, and the magnitude and direction of uncertainty are important aspects of any modeled map created. In addition, since there is a probability distribution built for each location of the map, the user can depict the error in different ways based on her/his goals and choice of risk (e.g., of overestimating or underestimating). For example, if in a study of the hemlock woolly adelgid (an insect associated with hemlock mortality), the cost of sending crews to a site with an insufficient amount of hemlock is greater than missing a site that might have sufficient hemlock to complete the study, i.e., the cost of overestimation is greater than that of underestimation. The user would thus want to choose a depiction of error that reflects the wish to avoid wasting field crew resources—one that reduces the risk of overestimation.

Like all techniques, however, there are also limitations to SGCS that make us want to pursue additional approaches. First, not all variables of interest have a strong spatial structure that can be modeled and used in SGCS to improve estimates. Second, satellite data and other relevant spatial data layers (e.g., topographic information, climatic information, soils information), are becoming increasingly available and affordable and are of increasing quality. Also, our technical ability to display and manipulate these data is continually improving. Finally, because of the sampling intensity of FIA data and the level of unexplained variation typically remaining, using only FIA data in the modeling limits the spatial resolution and levels of uncertainty in the final output.

To address the shortcomings of univariate geostatistical methods, we are incorporating this increasing wealth of additional ancillary environmental information into the modeling process. Many multivariate modeling techniques are available for this, each of which utilizes and maintains different characteristics of the sample data, has different output characteristics, and makes different assumptions. Many of these methods are being investigated throughout FIA (e.g., Lister and Hoppus 2002; McRoberts *et al.* 2002; Moeur and Riemann 1999; Moisen and Frescino 2002; Ohmann and Gregory 2002). The goal of our study was to investigate the use of multiple linear regression to make predictions of FIA attributes by answering the following questions:

- Does spatial structure matter when satellite-derived and ancillary spatial data are incorporated into the modeling of forest attributes?
- If we use a modeling technique such as multiple linear regression, how do we calculate or estimate the uncertainty?
- Using this technique, what are the characteristics of the output data set based on the original sample data and the ancillary data used?

Methods

Data.—In all, 141 100-percent forested FIA plots from the 1998 inventory were used from the study area in central and southern New Jersey. FIA defines a forested plot as being at least 1 acre in area and 120 feet wide, having a minimum of 10-percent stocking, and an undisturbed understory.[4] Variables calculated from plot data and chosen for modeling were: the relative basal area of pitch pine (pp%ba), relative basal area of red maple (rm%ba), and total

[3] Riemann, R. and Lister, A.J. Stochastic simulation for mapping ground inventory variables: Creating and using the species distribution maps. Unpublished report on file at USDA FS NE-FIA, 11 Campus Blvd., Newtown Square, PA 19073.

coniferous basal area on a plot (conifba). Because of their skewed distributions, all three variables were normal–score transformed before all analyses—a 1-1 linear transform to a perfectly normal distribution (Deutsch and Journel 1998).

Only forested plots were used because we wanted to model the characteristics of forest land rather than the distinction between forest and nonforest land. Nonforest land was applied afterward as a mask on the modeled output, using data derived from a classification dedicated to accurately identifying those classes (Zhu and Evans 1994). Separating these two tasks and focusing on modeling only the forested population was considered a valuable part of the modeling exercise. Mixed plots, those partly on forest land and partly on nonforest land, add an additional complication when attempting to match plot data with the 30-m pixel of Landsat Thematic Mapper (TM)-derived layers. With mixed plots, being off by even a half pixel in co-registration of the data sets can place the plot in an entirely different land use class, which is well below the georeferencing accuracies currently achievable. This did reduce the usable number of plots available (from 206 forested and mixed plots to the 141 completely forested plots). Future studies should explore ways to address this uncertainty and use this plot information, but in this study mixed plots were simply removed.

The following predictor variables were used:

- Spectral information derived from Landsat ETM+ (USGS Eros Data Center 2002): three tassel cap bands each from three seasons (spring, summer-leaf on, fall-leaf off), image dates: approx. 2000 (range: 1-255).
- Topographic variables derived from the 30-m digital elevation model (DEM): elevation, slope, aspect, position indicator (range: 0-100, representing location between the valley (0) and the ridgetop (100)).
- Soils variables derived from STATSGO (USDA 1993): soil quality, soil carbon, available water content (soil_awc) (range: 1-255).
- Spatial information – X, Y (converted to a range of 1-255).

Modeling Approach.—Descriptive statistics were calculated and plotted to assess the characteristics of the dependent variables, and the data were normal-score transformed (a 1-1 linear transform of the data to a perfectly normal distribution) before further analysis. Variography was performed and variogram descriptive statistics were calculated to assess the degree of spatial continuity of each of the dependent variables. Scatterplots were constructed and correlation statistics were produced to assess the degree of correlation with predictor data and eliminate predictor data layers that were not linearly related to the dependent data layers. Data redundancy was reduced by removing one variable from pairs of predictor variables that were collinear. We performed stepwise linear regression to make maps of predicted pp%ba, rm%ba and conifba, both including and excluding X and Y as possible predictor variables. To assess the accuracy of the regression model, we analyzed the model fit and performed a tenfold cross-validation procedure in which successive sets of 10 percent of the data were withheld from the model and subsequently predicted. Scatterplots of observed vs. actual values and residual plots were produced and assessed. Finally, we compared the characteristics of the output data sets with those of the original sample data, looking for differences that might be effects of the modeling technique or data sets used.

Results and Discussion

Characteristics of the Data.—The skewed distributions of the dependent variables indicated a probable need for transformation before analysis, and indeed a clearer spatial structure and stronger correlations with the independent variables were observed with the normal-score transformed data.

The following are results of variography and correlation analyses:

Species	Variogram nugget (% of variation explained)	r value
pp%ba	0.45 (55%)	0.4–0.6
rm%ba	0.8 (20%)	0.2–0.3
conifba	0.55 (45%)	0.5–0.7

Both pitch pine relative basal area (pp%ba) and total coniferous basal area (conifba) had a strong spatial structure and fairly strong correlations with the independent/explanatory variables. Red maple had weak spatial structure and weaker correlations with the independent/explanatory variables.

[4] U.S. Department of Agriculture, Forest Service. 2000. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.4. Unpublished report on file at USDA Forest Service, Forest Inventory and Analysis, Washington, DC.

Models.—The model developed by the linear regression using the full set of sample points and used for producing the final conifba map was:

conifba = -1.22186 + (0.03134*fall_wet) – (0.02951*fall_bright) + (0.02204* spr_green) - (0.01713 * sum_green) + (0.00265*y) - (0.00906*soil_awc)

$R^2 = 0.62$, p = 0.13, RMSE = 0.63.

The model used for the rm%ba final map was:

rm%ba = -1.1885 + (0.03983*sum_green) – (0.03552*fall_green) – (0.00774*position_ind) + (0.00825*soil_awc)

$R^2 = 0.25$, p = 0.046, RMSE = 0.88.

The model for predicting pp%ba was:

pp%ba = 1.63594 + (0.03333*spr_green) - (0.02823*sum_green) - (0.01128*soil_awc)

$R^2 = 0.54$, p = 0.0006, RMSE = 0.686.

The same larger set of potential predictor variables was provided for the development of the 10 validation models. Each of these models was similar to that developed from all the sample points.

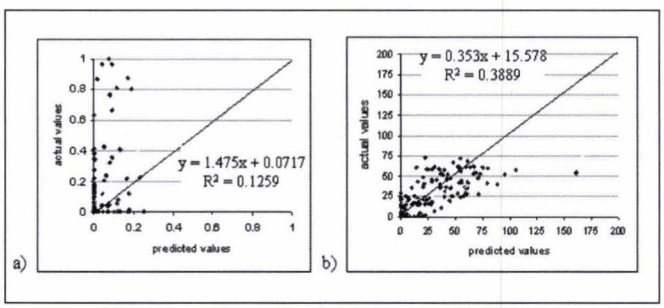
Calculating and Depicting Map Error/Uncertainty

How good is the model? How close to reality is the output map? How likely are we to find on the ground what is depicted in my map? A measure of uncertainty associated with each estimate can reveal some pertinent information, and it can capture one or more of the above elements of uncertainty since they are related. The most directly interpretable values for the user are measures of how the output map relates to reality, typically in terms of comparison to point locations on the ground. However, depending on the final use(s) of the output data set, other characteristics may also be important, such as the accuracy of area summaries or the spatial distribution of features.

We assessed the uncertainty of our regression models by examining the results of the tenfold cross validation. This involved dividing the data into 10 random parts, each containing 90 percent of the data, and running linear regression on each of the 10 sets. Each of the original values was then compared with that value predicted using the model created without it. Scatterplots of observed vs. predicted backtransformed values from this validation for each of the dependent variables are shown in figure 3. From these validation data we can calculate an RMSE value to describe the uncertainty associated with our output map. The validation RMSE with an average uncertainty of +/- 23 percent for estimates of red maple relative basal area, and 31 ft^2 for estimates of total coniferous basal area. However, RMSE is only a single average value for the entire map. How does the magnitude and direction of that difference vary with location and predicted value?

The previous analysis estimated the error associated with each known value. We also wanted to produce a spatial depiction of model uncertainty. Assuming that the validation we conducted using plots for which we had data provided a picture of the distribution of possible prediction errors, we grouped all the predicted values into classes, ensuring that enough plots fell in each, and calculated the average validation error associated with that class (table 2). We then reclassified our table of predictions to create an error map (fig. 4a) using table 2 as a lookup table. Figure 4b is the associated map of estimated values. Note that we are also assuming that we can associate a pixel (30- x 30-m grid cell) with a plot (a 4-point cluster of 1/24-acre circular plots spread over approximately 1 acre).

Figure 3.—*Actual vs. predicted values for a) proportion of red maple basal area and b) total coniferous basal area. Diagonal line is the line of 1:1 agreement.*



Checking Characteristics of the Output Data Set

What are the characteristics of the output maps in relation to the original sample data and the ancillary data used? When we checked the univariate statistics of the output data set, the predicted data duplicated the original data's sample histogram for all three variables fairly well (fig. 5). Looking at the correlogram/spatial structure of the output data set, the output is spatially more smoothed than the original data (table 3). This is partly due to the characteristics of the predictor data sets used.

Joint attribute structure—the relation between the forest variables predicted—is another important characteristic for those interested in querying the output data sets together. Although we did not model the three variables together, and thus did not explicitly attempt to retain this information in the modeling, it is nevertheless important to understand how this is expressed in the output data sets. Comparing the joint attribute structure in the output to the original data shows substantial similarities but also some truncation of the original ranges of values (fig. 6). Finally, looking at the output map and the input data sets that were used to create it, we observed that the characteristics of the source data sets were influencing the output in ways that may be undesirable. For example, in figure 7d, areas with the highest estimates (along the shorelines) are clearly influenced strongly by soil available water content (soil_awc) values (7a). In figure 7b, some relics of how the position_index value was calculated appear as "contour lines" in the rm%ba map. Such effects may be important clues to the driving factors associated with particular species, or, in this case, more likely suggest room for improvement in the model and/or in the

Table 2.—*Calculating average error for each class of predicted value—i.e., for each range of predicted values, the range of possible actual values*[a]

predicted classes	count	avg minus error (est too low)	avg plus error (est too high)	avg plus_minus error
0	25	3	0	2
10	17	6	3	5
20	19	4	9	6
30	24	12	15	13
40	11	13	15	14
50	16	18	13	16
60	17	16	19	17
80	12	21	23	22

[a] The average plus error can be sufficiently different from the average minus error such that one might want to depict them separately.

Figure 4.—*Map of the predicted values (a) and of the estimated uncertainty plotted from table 2 (b).*

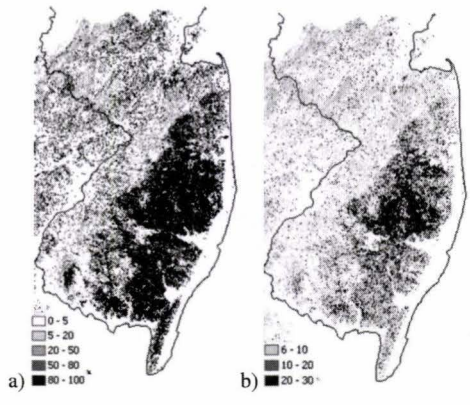


Figure 5.—*Univariate histograms of the sample data compared with predicted values.*

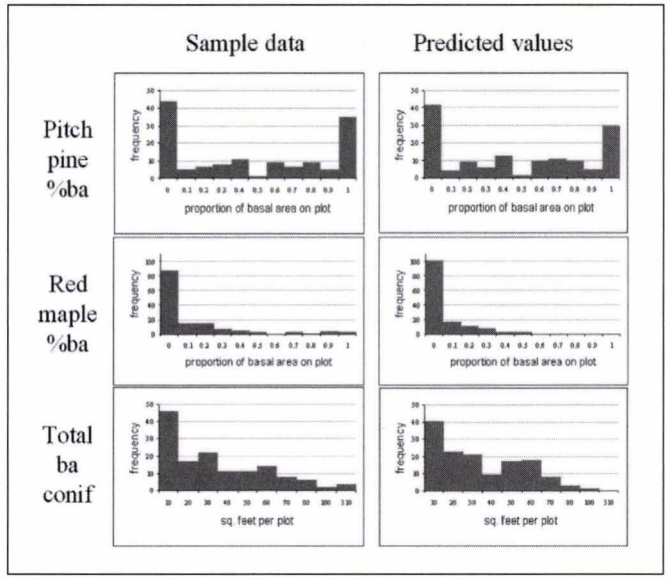
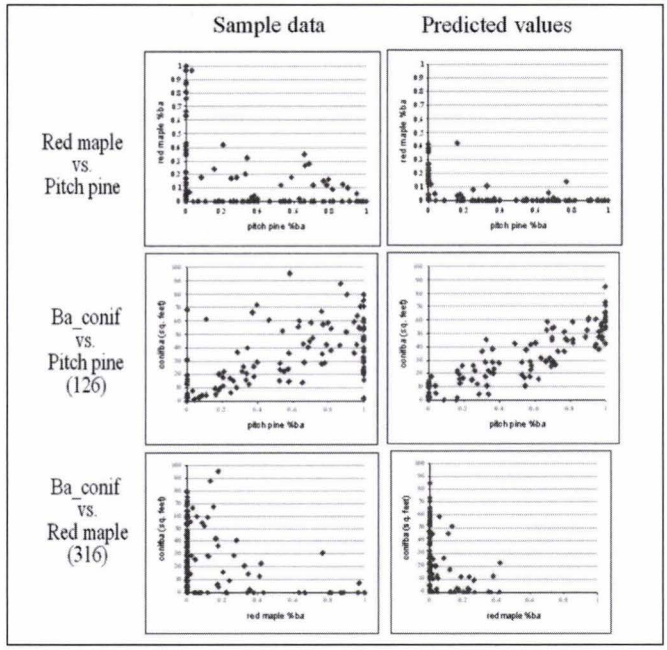


Figure 6.—*Scatterplots of the joint attribute structure of original and predicted data.*



independent/source data sets when they become available. The results also suggest that in future studies we may want to reevaluate the inclusion of ancillary spatial layers that adversely affect the final maps.

Incorporating Spatial Structure

On the basis of the nugget values calculated during the variography analysis, there is spatial structure in the sample data, particularly with relative pitch pine basal area and total coniferous basal area. In addition, the simple spatial location variables of standardized X and Y demonstrate some level of correlation (r values of 0.33 to 0.56 for X and 0.02 to 0.12 for Y)) with all three species or species-group variables. Whenever X or Y was one of the final independent variables in the model, it always smoothed the output data set (fig. 8), and did not noticeably change the amount of spatial structure remaining in the residuals (table 4). This limited effect of including or excluding X and/or Y as predictor variables in the model is probably because much spatial information is already implicitly contained in the satellite imagery and other spatial data sets when these are strongly correlated to the variable of interest. For variables that are both poorly correlated with the independent variables and contain a high level of spatial structure, incorporating spatial information could introduce important additional information (Goovaerts 1999, Moeur and Riemann 1999). None of the variables investigated contained this combination of characteristics. However, where including the spatial structure in the modeling is desirable, multivariate linear least

Table 3.—*Comparison of nugget values in the correlograms of the original values and the predicted values, expressed as % of the sill. A lower nugget indicates more spatial structure; there is a noticeable difference between the original and predicted values here.*

Item	Original	Predicted
rm%ba	80	12
conifba	55	8

Table 4.—*Nugget values in the residuals, expressed as % of the sill. A low nugget often indicates spatial structure not explained by the model. Including X in the model did not reduce the amount of spatial structure in the residuals.*

Item	Model	
	With X	Without X
rm%ba	52	68
conifba	100	100

Figure 7.—*Degree to which the spatial characteristics of the predictor data sets can contribute to the final map. This may or may not be realistic, but in the Northeast, where things are seldom driven so cleanly by a single variable, we would tend to suspect this as a characteristic of the data and model rather than of the phenomena: a) soil_awc; b) position_indicator; c) plus/minus error associated with predicted estimate; d) predicted %ba value for red maple.*

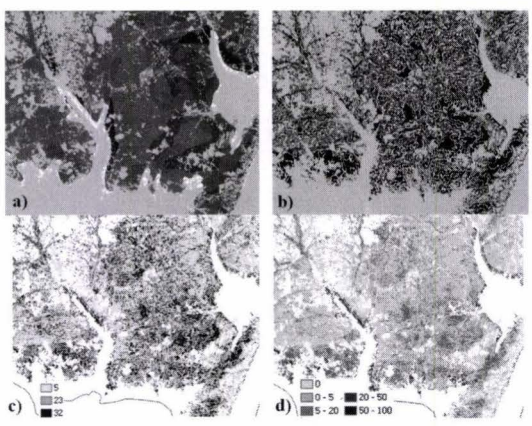
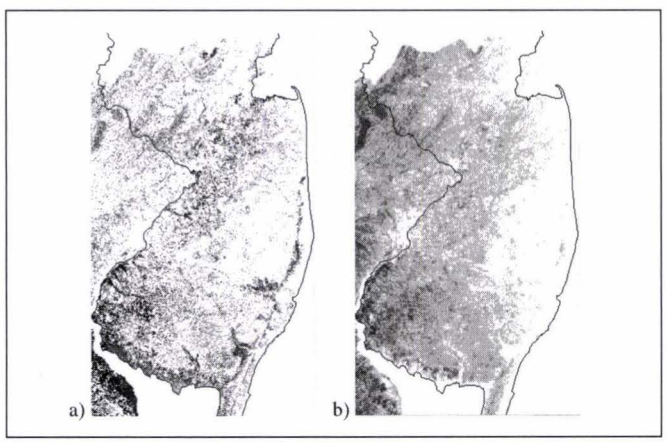


Figure 8.—*Effect of using X in the model for estimating red maple relative basal area (rm%ba): a) a map of the model developed excluding XY as optional predictor layers (-0.73814 + (0.01412 * sum_green) – (0.00759 * position_ind) – (0.00593 * X)), b) a map of the model developed including XY as options (-1.18850 + (0.03983 * sum_green) – (0.03552 * fall_green) + (0.00825 * soil_awc) – (0.00774 * position_ind)). Note the increased smoothing of the results in b) when the smoothly varying variable X is included.*



squares regression is probably not the best tool to take advantage of spatial information. Because it is mostly a per pixel modeling technique, linear regression does not easily incorporate information on the value, distance, or direction of neighboring data when producing an estimate. Other approaches may be much more effective in taking advantage of this spatial information.

Conclusions

Know the characteristics of your data and the phenomena going in—it will help greatly in the modeling, both in effectiveness and accuracy, for each has different characteristics that can affect the effectiveness of a particular modeling technique. Similarly, check the characteristics of the data coming out—it will help in understanding the characteristics of the output data set, which, in turn, may result in iterative improvement even if it does not contribute directly to a calculation of error/uncertainty. As stated earlier, the characteristics of the spatial data sets used in predicting the variable of interest affected the characteristics of the output. Such effects may be important clues to the driving factors associated with particular species, or, more likely in this case, signal possible improvements in the model and/or in the independent/source data sets.

Having some measure of uncertainty is SO important! Each of these maps is only a modeled estimate of what is occurring on the ground, so there is *always* some level of uncertainty as to the degree to which the modeled map reflects reality. Providing a measure of uncertainty with each estimate gives the user additional information to work with. With linear regression, a single validation RMSE value, such as a single %accuracy value for a classified map, doesn't tell the whole story with the mapped output of a regression model because we know there is spatial and class variability to that error and we want to know where that occurs. A single RMSE value calculated without the use of a validation data set is even less satisfactory because an uncertainty value calculated only from data that went into the modeling does not account for errors in our admittedly less-than-perfect input data sets. Error/uncertainty from comparisons between modeled estimates and plot data can be calculated when both are of a reasonably similar scale/resolution (e.g., a 30-m pixel with a 1/6-acre clustered plot).

Does spatial structure matter when we incorporate satellite-derived and ancillary spatial data sets into the modeling of forest attributes; that is, do we lose potential information by not including it? In this study, bringing the satellite and other layers into the modeling seemed to account for much of the variability that the spatial structure was describing. However, if we wish to include spatial structure in the modeling (as might be the case with a variable with weak correlations with the independent variables but a strong spatial structure), bringing X and Y into the regression as simple variables is only a partial solution with some consequences (i.e., smoothing), and is probably not the best technique for this task.

There are many characteristics of a spatial data set. Many maps are a compromise of some characteristics in favor of others, e.g., smoothing the map to discern patterns at the expense of local heterogeneity. Which aspects are most important in the output map will depend on how the data will be used. What the maps are being used for will direct/dictate how we look at them, what we consider to be accurate (or the most important aspect of accuracy), what we consider to be the dominant characteristic that makes us accept or reject a map, and what modeling techniques we choose because of the characteristics they preserve or the characteristics of their output.

The models produced from linear regression procedures in this study are by no means the best that could be obtained given the independent data currently available. However, they point out potential characteristics and tendencies that can result when input variables are used in a regression model to simply predict forest attributes by their relationship to other spatial variables. Characteristics of the input data sets, which may be derived from numerous sources and via other modeling techniques, can greatly influence the characteristics of the output.

Literature Cited

Deutsch, C.V.; Journel, A.G. 1998. GSLIB: geostatistical software library and user's guide. 2nd ed. New York: Oxford University Press.

Goovaerts, P. 1999. Performance comparison of geostatistical algorithms for incorporating elevation into the mapping of precipitation. In: Proceedings, IV international conference on geocomputation; 1999 July; Fredericksburg, VA. 17 p. www.geovista.psu.edu/sites/geocomp99/Gc99/023/gc_023.htm. [31 January 2003]

Lister, A.J.; Hoppus, M. 2003. Small area estimation using FIA plots and satellite imagery. In: Proceedings, 2002 IUFRO seminar on statistics and information technology in forestry; September 8–12; Blacksburg, VA. Blacksburg, VA: Department of Forestry, Virginia Polytechnical Institute and State University: 45–52.

McRoberts, R.E.; Nelson, M.D.; Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the *k*-nearest neighbors technique. Remote Sensing of the Environment. 82(2/3): 457–468.

Moeur, M.; Riemann, R.R. 1999. Preserving spatial and attribute correlation in the interpolation of forest inventory data. In: Lowell, K; Jaton, A., eds. Spatial accuracy assessment: land information uncertainty in natural resources. Chelsea, MI: Ann Arbor Press: 419–430.

Moisen, G.G.; Frescino, T.S. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling. 157: 209–225.

Ohmann, J.L.; Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, USA. Canadian Journal of Forest Research. 32: 725–741.

Riemann, R.H.; Ramirez, M.A.; Drake, D.A. 1997. Using geostatistical techniques to map the distribution of tree species from ground inventory data. In: Gregoire, T.; Brillinger, D.R.; Diggle, P.J. *et al.*, eds. Modelling longitudinal and spatially correlated data: methods, applications, and future directions. Lecture notes in statistics, vol. 122. New York, NY: Springer-Verlag: 187–198.

Rossi, R.E.; Borth, P.W.; Tollefson, J.J. 1993. Stochastic simulation for characterizing ecological spatial patterns and appraising risk. Ecological Applications. 3(4): 719–735.

U.S. Department of Agriculture. 1993. State soil geographic data base (STATSGO) data user's guide. Misc. Publ. 1492. Washington, DC: U.S. Department of Agriculture, Soil Conservation Service.

Zhu, Z.; Evans. D.L. 1994. U.S. forest types and predicted percent forest cover from AVHRR data. Photogrammetric Engineering & Remote Sensing. 60(5): 525–531.

Development and Validation of Spatially Explicit Habitat Models for Cavity-nesting Birds in Fishlake National Forest, Utah

Randall J. Schultz, Jr.[1]; Thomas C. Edwards, Jr.[2]; Gretchen G. Moisen[3]; and Tracey S. Frescino[4]

Abstract.—The ability of USDA Forest Service Forest Inventory and Analysis (FIA) generated spatial products to increase the predictive accuracy of spatially explicit, macroscale habitat models was examined for nest-site selection by cavity-nesting birds in Fishlake National Forest, Utah. One FIA-derived variable (percent basal area of aspen trees) was significant in the habitat model; however, the incorporation of FIA stand structure information did not increase model accuracy. Cavity-nesting birds respond strongly to nest-tree attributes unable to be modeled spatially for this study. Future modeling efforts should focus on larger taxa (e.g., ungulates) and richness/diversity studies.

Background

Recent efforts in wildlife habitat modeling have focused developing spatially explicit habitat models (Carroll *et al.* 1999, Dettmers and Bart 1999, Edwards *et al.* 1996, Knick and Rotenberry 1995, Lawler and Edwards 2002, Mitchell *et al.* 2001, Reunanen *et al.* 2002). The ability to build spatially explicit habitat models is desirable for several reasons. First, the models can be used to make spatial predictions across large and remote regions. Second, they often rely on remotely sensed data and/or pre-existing habitat data. These data may be quickly and easily applied to habitat modeling. Field habitat data collection, however, may often be time-consuming and labor intensive (Mitchell *et al.* 2001).

Most spatially explicit habitat models use cover-type information, or macroscale information, to predict species presence (Edwards *et al.* 1996, Lawler and Edwards 2002, Reunanen *et al.* 2002, among numerous others). Despite its ease of use, coarse-scale cover-type information may be too general and limited for predicting species reliant on the structure and condition of individual trees or stands (Lawler 1999, Lawler and Edwards 2002, Schultz and Edwards, unpublished data). Thus, ecologists have begun to incorporate finer-scale forest structural variables (i.e., stand structure) into spatially explicit habitat models (Carroll *et al.* 1999, Reunanen *et al.* 2002).

To incorporate forest structure variables, ecologists are searching for methods of modeling forest structure across space (Frescino *et al.* 2001, Moisen and Edwards 1999, Moisen and Frescino 2002). One technique involves converting statistical models of forest structure to spatially explicit maps of forest attributes (e.g., basal area, snag density, live trees per acre, canopy height, biomass, etc.) (Frescino *et al.* 2001, Frescino and Moisen 2004, Terletzky and Frescino 2004). Pre-existing USDA Forest Service Forest Inventory and Analysis (FIA) field data are used as response variables, and a combination of environmental variables and remotely sensed data are used as predictor variables. The resulting models are converted to spatially explicit prediction maps, and the mapped variables can then be used in wildlife habitat modeling.

The primary objective of this research was to determine whether incorporating FIA-generated spatial products (hereafter mesoscale) improved the predictive accuracy of macroscale habitat models for cavity-nesting bird nests in Fishlake National Forest, Utah. The results were then used to assess both the utility of FIA-generated spatial products in habitat modeling for cavity-nesting birds, and the current abili-

[1] Graduate Research Assistant, Department of Forest, Range, and Wildlife Sciences, and Ecology Center, Utah State University, Logan, UT 84322–5210. Phone: 410–382–7806; fax: 435–797–4025; e-mail: randalljs@cc.usu.edu.

[2] Assistant Leader and Associate Professor, USGS Biological Resources Division, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322–5210. Phone: 435–797–2529; fax: 435–797–4025; e-mail: tce@nr.usu.edu.

[3] Research Forester, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401. Phone: 801–625–5384; fax: 801–625–572; e-mail: gmoisen@fs.fed.us.

[4] Forester, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401. Phone: 801–625–5402; fax: 801–625–5723; e-mail: tfrescino@fs.fed.us.

ty of spatially explicit models to predict the presence of cavity-nesting bird nests in Fishlake National Forest, Utah.

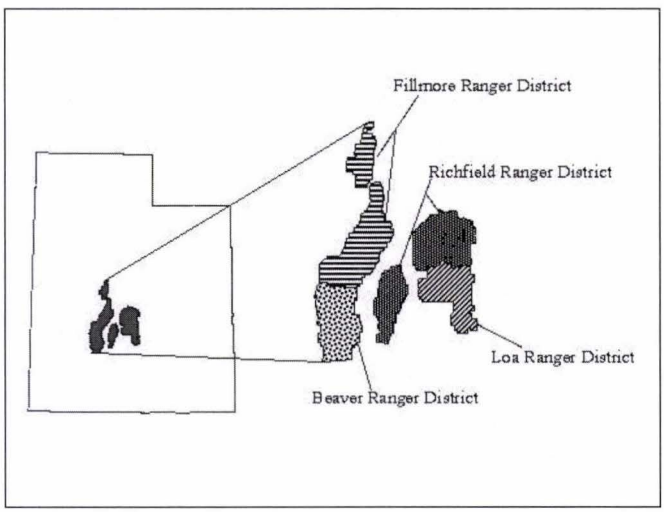
Methods

Study Area

The study area was the Fishlake National Forest, located in southern Utah at the southern end of the Wasatch Mountains (fig. 1). The study area encompassed sections of four ranger districts (Richfield, Loa, Fillmore, and Beaver) across three general mountain areas (fig. 1). This region of Utah is characterized by high mountains (~2,000 m to ~ 4,000 m) consisting of broad, rolling plateaus, large alpine meadows, and large areas of aspen (*Populus tremuloides*) forest.

Vegetation at low-elevation sites on the study area consists primarily of aspen stands interspersed with sagebrush meadows, ponderosa pine (*Pinus ponderosa*), curl-leaf mahogany (*Cercocarpus ledifolius*), gambel oak (*Quercus gambelii*), Utah juniper (*Juniperus osteosperma*) and pinyon pine (*Pinus edulis*). The vegetation at middle to high elevations consists of an aspen/mixed-conifer (Douglas-fir [*Pseudotsuga menziesii*]; Engelmann spruce [*Picea engelmannii*]; white fir [*Abies concolor*]; subalpine fir [*Abies lasiocarpa*]), meadow matrix. The vegetation grades into a spruce-fir forest until upper treeline.

Figure 1.—*The Fishlake National Forest in southern Utah, including the location of ranger districts.*



Study Species

The study species included all cavity-nesting birds found to nest in aspen communities of the forest. The species included six primary cavity-nesting birds: red-naped sapsucker (*Sphyrapicus nuchalis*), northern flicker (*Colaptes auratus*), hairy woodpecker (*Picoides villosus*), downy woodpecker (*Picoides pubescens*), three-toed woodpecker (*Picoides tridactylus*), and red-breasted nuthatch (*Sitta canadensis*); and six secondary cavity-nesting birds: tree swallow (*Tachycineta bicolor*), violet-green swallow (*Tachycineta thalassina*), mountain chickadee (*Poecile gambeli*), mountain bluebird (*Sialia currucoides*), western bluebird (*Sialia mexicana*), and house wren (*Troglodytes aedon*).

Study Design

We built habitat models based on presence/absence data for nests of cavity-nesting birds. To determine if the addition of mesoscale variables improved macroscale model accuracy, we built and validated predictive models using only macroscale variables and additional multiscale models using both macroscale variables and mesoscale variables. Model building data were collected in 2001, and validation data were collected in 2002. We compared model performance using the percent correctly classified (PCC), sensitivity, specificity, and the area under curve (AUC) values.

Nest Searches

Sample locations were identified using a 30-m resolution digital vegetation map from the Utah Gap Analysis Project (Edwards *et al.* 1998, Homer *et al.* 1997). Sample locations were restricted to aspen stands adjacent to meadow and/or conifer cover types. A total of 14 locations were searched during the study. We selected nine locations for model building during the summer of 2001, and reserved five locations for model validation during the summer of 2002. All the model-building locations were located on the Richfield Ranger District and a small section of the Loa Ranger District (fig. 1). To select validation locations for 2002, we stratified the forest geographically and reserved new locations in previously unsearched sections of the national forest. Thus, the 2002 validation locations were located on the Fillmore and Beaver Ranger Districts, and another section of the Loa Ranger District (fig. 1).

We systematically surveyed study locations for active nests of cavity-nesting birds from late May until early July. We considered a nest active if it showed evidence of incubation, presence of eggs, presence of young, and/or feeding activity. To mark the active nests, we recorded the UTM coordinates at each nest using a global positioning system. Several non-nest locations were selected at the end of each breeding season. We considered a non-nest location to be an aspen tree (>10 cm d.b.h. and >1.4 m high) within a previously searched location. We randomly selected non-nest locations that were 100-150 m apart from each other and each active bird nest.

Table 1.—*Habitat variables and their descriptions, Fishlake National Forest, Utah (macroscale variables based on cover-type metrics measured at a 15 ha scale; mesoscale variables obtained from FIA-generated spatial products and measured at a 30-m scale)*

	Variable name	Description
Macroscale	%open	Percent landscape of open land, including meadows
	%aspen	Percent landscape of aspen forest
	%conifer	Percent landscape of conifer forest
	%mixed	Percent landscape of mixed conifer/aspen forest
	Lpopen	Largest patch of open land (% of landscape)
	Lpaspen	Largest patch of aspen forest (% of landscape)
	Edaspen	Edge density of aspen forest (m/ha)
	Pr	Patch richness of the landscape (#)
	Sdi	Simpson's Landscape Diversity Index (%)
Mesoscale	Ba	Live tree basal area (sq ft/acre)
	Crcov	Crown cover (%)
	Stage	Stand age (yrs)
	Tpa	Live trees per acre (trees/acre)
	Vol	Net volume of trees (cu. ft./acre)
	Qmd	Quadratic mean diameter of trees (in)
	Bio	Live tree biomass (tons/acre)
	Aspba	Aspen basal area (%)
	Asprot	Aspen rot (presence/absence)
	Snags	Number of snags
	Avtrht	Average tree height (ft)

Habitat Data Collection

Based on prior statistical analysis, we chose 15 ha as the macroscale for cavity-nesting birds in Fishlake National Forest (Schultz and Edwards, unpublished data). This scale approximates the home-range of the northern flicker, the largest and most abundant bird in the data set (Dunning 1993, Lawrence 1967). All macroscale variables were measured at this scale.

All macroscale variables were generated from 30-m resolution vegetation data layers in Arc/INFO GIS. The vegetation data layers were derived from the 1999 National Land Cover Data set, which was created using Landsat Thematic Mapper imagery and ancillary data (Vogelmann *et al.* 2001). Five general cover types were considered: open land (shrublands, grasslands, wetlands), aspen forest, conifer forest, mixed forest, and an "other" cover type. Using a square moving window centered on each nest and non-nest, we estimated landscape attributes using FRAGSTATS (McGarigal and Marks 1995). We selected nine attributes we felt were relevant to cavity-nesting bird habitat, including the percent landscape of cover types, edge density of aspen, and richness/diversity measurements (table 1).

We used a 30-m pixel to represent the mesoscale, or stand habitat. This scale was the smallest measurement possible in this study. In addition, this scale roughly approximates the size of a 0.04 ha plot, a commonly used field measurement in avian habitat studies (James and Shugart 1970, Noon 1981). Mesoscale measurements were derived from 30-m resolution digital maps of FIA-derived variables, including aspen basal area, number of snags, number of live trees per acre, and canopy height (table 1).

The FIA data were modeled spatially using several different statistical tools, including generalized additive models (GAMs) and multivariate adaptive regression splines (MARS). The models were then converted to spatially explicit prediction maps of mesoscale forest structure, which were then used in habitat modeling (Terletzky and Frescino 2004).

Table 2.—*Estimates of model fit for the stepwise logistic regression habitat models of cavity-nesting bird nesting habitat in Fishlake National Forest, Utah (variables significant at the p=0.05 statistical level)*

Model / variable	R^2	D	Estimate	Standard error
Macroscale	0.063	0.341		
Intercept			-0.406	0.216
%open			0.025	0.006
Multiscale	0.077	0.365		
Intercept			-0.773	0.287
%open			0.029	0.006
Aspba			0.008	0.004

Statistical Models

To reduce redundancy of information in the habitat models, we examined correlations among variables and retained variables we deemed to have high ecological relevance. We chose a Pearson's correlation coefficient of 0.7 to be the minimum value necessary for variable elimination. We used stepwise logistic regression (Hosmer and Lemeshow 1989, SAS version 8) to model the presence of cavity-nesting birds based on habitat associations.

To assess the relative ability of the macroscale and multiscale habitat models to predict nest presence, we searched the five validation locations during the summer of 2002 and observed how well the 2001 models predicted nests and non-nests. We assessed model performance using various measures of model classification accuracy and performance, including percent correctly classified (PCC), sensitivity (true positive fraction), specificity (true negative fraction), and the threshold-independent area under curve (AUC) value from receiver operating characteristic (ROC) analysis (Fielding and Bell 1997, Zweig and Campbell 1993). We used a 0.5 decision threshold for all threshold-dependent classification analyses.

Results

Model Development

We found a total of 227 nests during the course of this study: 165 nests for model building (2001) and 62 nests for model validation (2002). In addition, we selected 170 non-nest locations: 117 for model building and 53 for model validation.

Table 3.—*Relative model performance of the macroscale and multiscale models of cavity-nesting bird nesting habitat in Fishlake National Forest, Utah (PCC, sensitivity, and specificity values based on a 0.5 classification threshold)*

Model	PCC (%)	Sensitivity	Specificity	AUC
Macroscale	63.5	0.774	0.472	0.670
Multiscale	57.4	0.677	0.453	0.680

Cavity-nesting birds increased with the percent of open habitat in both the macroscale and multiscale models (table 2). In the multiscale models, cavity-nesting birds also increased with the percent basal area of aspen. Model fit based on R^2 and Somer's D statistic was low for both models, and fit differed only marginally between the models (table 2).

Model Validation

In general, incorporating mesoscale FIA-derived information did not increase the accuracy of spatially explicit habitat models for cavity-nesting birds (table 3). Overall, classification accuracy was generally poor, with the macroscale model predicting marginally better than the multiscale models (table 3). Sensitivity values were remarkably higher than their corresponding specificity values, suggesting both models tended to overpredict bird habitat. Specificity values were low for both models. AUC values did not differ much between models (table 3).

Discussion

The results of this study suggest that mesoscale FIA-derived information can be applied to wildlife habitat modeling. The positive association between nest presence and aspen basal area supports this conclusion. Although spatially explicit FIA information can be used in habitat modeling, it did not increase the ability to predict nest presence of cavity-nesting birds in this study.

Two factors may account for the inability of mesoscale FIA-derived information to increase model accuracy. First, scale is inevitably an issue of concern in ecology (Levin 1992, Wiens 1989). A 30-m resolution may be too coarse a scale to predict nesting habitat for cavity-nesting birds. The distribution of cavity-nesting bird nests might better be predicted by nest tree attributes and stand structure in areas much smaller than 30 m. Cavity-nesting birds are strongly associated with nest tree attributes, including tree diameter and the evidence of decay (e.g., fungal conks) (Conner *et al.* 1976, Daily 1993, Daily *et al.* 1993, Dobkin *et al.* 1995, Kilham 1971, Lawler 1999). Fungal conks indicate heartrot, which facilitates excavation by cavity-nesting birds. However, the presence of fungal conks is a variable for which we cannot currently build spatially explicit maps. Future habitat modeling efforts for cavity-nesting birds in this and other similar regions should focus on finding macroscale and mesoscale surrogates for fungal conks and/or heartrot.

Second, a habitat model is only as accurate as the data used to build the model. Map error is a concern, and both vegetation modeling error and spatial error may have influenced the accuracy of the habitat models. Future vegetation mapping should focus on more accurate maps of forest structure and rigorous field-validation.

In aspen forests of Fishlake National Forest, ecologists cannot currently predict nest presence of cavity-nesting birds accurately without field habitat data. FIA-generated spatial products may have more utility for other species and issues than cavity-nesting bird nest-site selection. These products may be useful for ungulates and other large animals, where 30-m resolution may be more appropriate. Species richness and diversity studies may also benefit from this information. Future efforts concerning the utility of FIA-generated spatial products in wildlife habitat modeling should continue on these fronts.

Acknowledgment

We would like to extend thanks to Pat Terletzky for assistance with the computer applications in this research. We thank Fishlake National Forest, especially Kreig Rasmussen and Bob Campbell, for making the field component possible. This work would not have been possible without the work of field technicians Darlene Kilpatrick and Rita-Lyle Reitzel.

This project was funded by the USDA Forest Service, Rocky Mountain Research Station in Ogden, Utah, in cooperation with the USGS Biological Resources Division, Utah Cooperative Fish and Wildlife Research Unit, the Fishlake National Forest, and the Utah State University Ecology Center.

Literature Cited

Carroll, C.; Zielinski, W.J.; Noss, R.F. 1999. Using presence-absence data to build and test spatial habitat models for the fisher in the Klamath Region, U.S.A. Conservation Biology. 13(6): 1344–1359.

Conner, R.N.; Miller, O.K.; Adkisson, C.S. 1976. Woodpecker dependence on trees infected by fungal heart rots. Wilson Bulletin. 88(4): 575–581.

Daily, G.C. 1993. Heartwood decay and vertical distribution of Red-naped Sapsucker nest cavities. Wilson Bulletin. 105(4): 674–679.

Daily, G.D.; Ehrlich, P.R.; Haddad, N.M. 1993. Double keystone bird in a keystone species complex. Proceedings of the National Academy of Sciences. 90: 592–594.

Dettmers, R.; Bart, J. 1999. A GIS modeling method applied to predicting forest songbird habitat. Ecological Applications. 9(1): 152–163.

Dobkin, D.S.; Rich, A.C.; Pretare, J.A.; Pyle, W.H. 1995. Nest-site relationships among cavity-nesting birds of riparian and snowpocket aspen woodlands in the northwestern Great Basin. The Condor. 97: 694–707.

Dunning, J.B., Jr. 1993. CRC handbook of avian body masses. Boca Raton, FL: CRC Press, Inc. 371 p.

Edwards, T.C., Jr.; Deshler, E.T.; Foster, D.; Moisen, G.G. 1996. Adequacy of wildlife-habitat relation models for estimating spatial distributions of terrestrial vertebrates. Conservation Biology. 10: 263–270.

Edwards, T.C., Jr.; Moisen, G.G.; Cutler, D.R. 1998. Assessing map uncertainty in remotely sensed, ecoregion-scale covermaps. Remote Sensing and Environment. 63: 73–83.

Fielding, A.H.; Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation. 24(1): 38–49.

Frescino, T.S.; Edwards, T.C., Jr.; Moisen, G.G. 2001. Modelling spatially explicit forest structural attributes using generalized additive models. Journal of Vegetation Science. 12: 15–26.

Frescino, T.S.; Moisen, G.G. 2004. Predictive mapping of forest attributes on the Fishlake National Forest. In: McRoberts, R.E.; *et al.*, eds. Proceedings of the 4th annual Forest Inventory and Analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC–252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Homer, C.H.; Ramsey, R.D.; Edwards, T.C., Jr.; Falconer, A. 1997. Landscape cover-type modeling using a multi-scene TM mosaic. Photogrammetric Engineering and Remote Sensing. 63: 59–67.

Hosmer, D.W.; Lemeshow, S. 1989. Applied logistic regression. New York, NY: John Wiley and Sons. 307 p.

James, F.C.; Shugart, H.H. 1970. A quantitative method of habitat description. Audubon Field Notes. 24(6): 727–736.

Kilham, L. 1971. Reproductive behavior of Yellow-bellied Sapsuckers. Part I. Preference for nesting in *Fommes*-infected aspens and nest hole interrelations with flying squirrels, raccoons, and other animals. Wilson Bulletin. 83: 159–171.

Knick, S.T.; Rotenberry, J.T. 1995. Landscape characteristics of fragmented shrubsteppe habitats and breeding passerine birds. Conservation Biology. 9(5): 1059–1071.

Lawler, J.J. 1999. Modeling habitat attributes of cavity-nesting birds in the Uinta Mountains, Utah: a hierarchical approach. Logan, UT: Utah State University. Ph.D. dissertation.

Lawler, J.J.; Edwards, T.C., Jr. 2002. Landscape patterns as predictors of nesting habitat: a test using four species of cavity-nesting birds. Landscape Ecology. 17: 233–245.

Lawrence, L. de K. 1967. A comparative life-history study of four species of woodpeckers. Ornithological Monographs 5. 156 p.

Levin, S.A. 1992. The problem of pattern and scale in ecology. Ecology. 73(6): 1943–1967.

McGarigal, K.; Marks, B.J. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-351. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 122 p.

Mitchell, M.S.; Lancia, R.A.; Gerwin, J.A. 2001. Using landscape-level data to predict the distribution of birds on a managed forest: effects of scale. Ecological Applications. 11(6): 1692–1708.

Moisen, G.G.; Frescino, T.S. 2002. Comparing five modeling techniques for predicting forest characteristics. Ecological Modeling. 157: 209–225.

Moisen, G.G.; Edwards, T.C., Jr. 1999. Use of generalized linear models and digital data in a forest inventory of Northern Utah. Journal of Agricultural, Biological, and Environmental Statistics. 4: 372–390.

Noon, B.R. 1981. Techniques for sampling avian habitats. In: Capen, D.E., ed. The use of multivariate statistics in studies of wildlife habitat. Gen. Tech. Rep. RM-87. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 45–52.

Reunanen, P.; Nikula, A.; Monkkonen, M.; *et al.* 2002. Predicting occupancy for the Siberian Flying Squirrel in old-growth forest patches. Ecological Applications. 12(4): 1188–1198.

Terletzky, P.; Frescino, T. 2004. Integrating spatial components into FIA models of forest resources: some technical aspects. In: McRoberts, R.E.; *et al.* eds. Proceedings of the 4th Annual Forest Inventory and Analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC–252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completing of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing. 67: 650–662.

Wiens, J.A. 1989. Spatial scaling in ecology. Functional Ecology. 3: 385–397.

Zweig, M.H.; Campbell, G. 1993. Receiver-operating characterisic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry. 39(4): 561–577.

Integrating Spatial Components into FIA Models of Forest Resources: Some Technical Aspects

Pat Terletzky[1] and Tracey Frescino[2]

Abstract.—We examined two software packages to determine their feasibility of implementing spatially explicit, forest resource models that integrate Forest Inventory and Analysis data (FIA). ARCINFO and Interactive Data Language (IDL) were examined for their input requirements, speed of processing, storage requirements, and flexibility of implementing. Implementations of two models was compared across three mapping extents in the two software packages. IDL completed the models in approximately half the time that ARCINFO did and required less memory.

The Rocky Mountain Research Station (RMRS), Interior West Forest Inventory and Analysis (IWFIA) program, Ogden, Utah, developed spatially explicit forest resource models that integrated forest inventory data with spatial data (satellite imagery and elevation data) (Frescino and Moisen 2004). Model generation was done with the Multivariate Adaptive Regression Splines program (MARS; Friedman 1991), a flexible, nonparametric regression-modeling tool that establishes relationships between a forest resource attribute and spatial information.

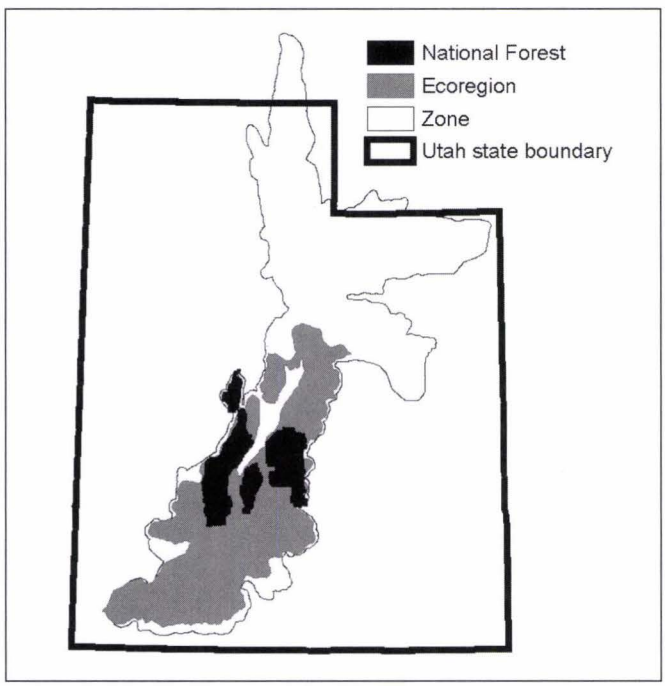
We were provided the MARS models and asked to create a user-friendly method of implementing the models to generate spatially explicit forest resource maps. Incorporating the spatial information into the forest resource models required a large amount of memory, special computing capabilities, and fast computers. We evaluated two software packages, ARCINFO (ESRI Inc., Redlands, California) and Interactive Data Language (IDL, Research Systems, Inc.; www.rsinc.com) based on four parameters: required inputs, processing speed, required memory, and ease of implementation. Although IDL is not common in the Forest Service (FS), we decided to examine it after several attempts with ARCINFO failed.

Methods

We implemented two MARS models in ARCINFO and IDL, and compared the time it took to complete the models and the storage required. The simple model had 6 input files, 13 intermediate files, and 1 output file; the complex model had 9 input files, 20 intermediate files, and 1 output file. Each model was implemented across three mapping extents (fig. 1). The largest mapping extent was a zone level, approximately 6,997,000 hectares, and the smallest extent was a forest level, approximately 627,000 hectares. The ecoregion level (Bailey 1980) was between the zone and forest levels and was 2,925,000 hectares (fig. 1).

We addressed two issues before implementing the models: (1) conversion of the spatially explicit models into the specific language for each software and (2) input and output data for-

Figure 1.—*Depiction of the three mapping scales used to examine two spatially explicit forest resource models.*



[1] Utah State University, Logan, UT.
[2] USDA FS FIA, IWFIA, Ogden, UT.

Figure 2.—*Conversion of a spatially explicit MARS model (a) in ASCII text to ARCINFO AML format (b) based on the Visual Basic conversion program. The original model and the AML output contain references to grids (Veg, Elev, Lat, and Long) and intermediate products to generate the final output (BF2, BF4, BF5, and BF6).*

(a) Forest Resource Model (Mars)

```
BF2 = max(44 – Veg, 0)
BF4 = max(1380 – Elev, 0) x BF2
BF5 = max(Lat – 1800, 0)
BF6 = max(12336 – Long, 0) x BF4

Final_Grid = (0.992 + 0.987 x BF2 + .002 x BF4 - .032 x BF5 + .266 x BF6)
```

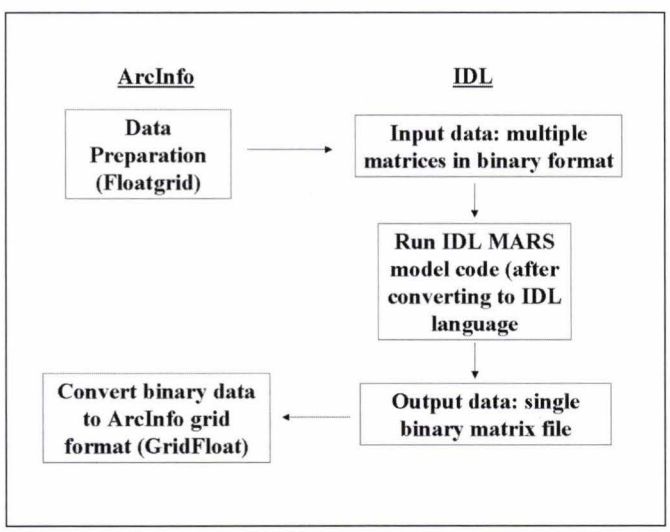
(b) Equivalent AML code

```
BF2 = con((44 – Veg >= 0), 44-Veg, 0)
BF4 = con((1380 – Elev >= 0), 1380 – Elev, 0) x BF2
BF5 = con((Lat – 1800 >= 0), Lat-1800, 0)
BF6 = con((12336 – Long >= 0) 12336-Long, 0) x BF4

Final_Grid = (0.992 + 0.987 x BF2 + .002 x BF4 - .032 x BF5 + .266 x BF6)
```

mats. The models were originally in ASCII format and required conversion to the specific language for each software package. Conversion into the ARCINFO software language (ARC Macro Language – AML) was done by a Visual Basic program, written by John Nelson, Ogden IWFIA (fig. 2). This program required user input as to file locations and file output names, and resulted in an AML code, which was then run in ARCINFO. Using this conversion program reduced the number of user input errors caused by typing the model into the computer by hand. Since there was no conversion program to create the corresponding IDL program, the MARS model was transcribed to the IDL format by hand typing the correct IDL code.

Figure 3.—*Process of implementing a spatially explicit model in IDL.*


ArcInfo
IDL
Data Preparation (Floatgrid)
Input data: multiple matrices in binary format
Run IDL MARS model code (after converting to IDL language
Output data: single binary matrix file
Convert binary data to ArcInfo grid format (GridFloat)

The original input and final output data format was an ARCINFO grid. Grids are a specific type of raster data unique to ARCINFO. Raster data are simply a series of columns and rows with each cell having a value that represents a specific class (e.g., 1 = water, 2 = agriculture, 3 = urban), or a single value from continuous data (e.g., elevation, soil pH). We required ARCINFO grids as the original inputs and final outputs because it is an industry standard and many Forest Service Geographical Information Systems (GIS) facilities have the ARCINFO software. Whereas ARCINFO required no changes in input or output data, IDL required converting the input grids to binary format, and the output binary files to grids, which required ARCINFO (fig. 3).

Figure 4.—*Comparison of the time it took to run a simple mode (6 inputs, 13 intermediate grids (binary data for IDL), and one output) for three levels in the ARCINFO and the IDL software packages.*

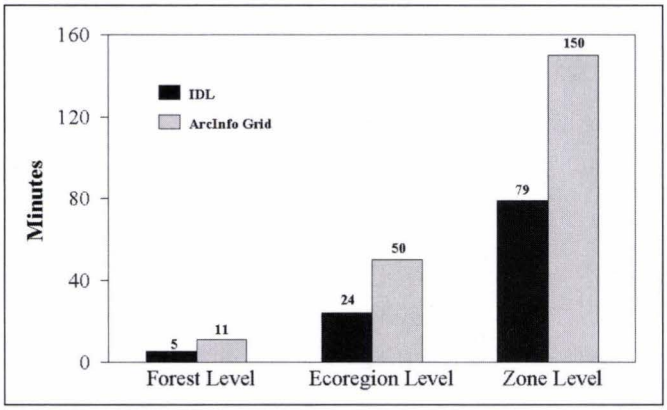


Figure 5.—*Comparison of the time it took to run a complex model (9 inputs, 20 intermediate grids (binary data for IDL) and one output) for three levels in the ARCINFO and the IDL software packages.*

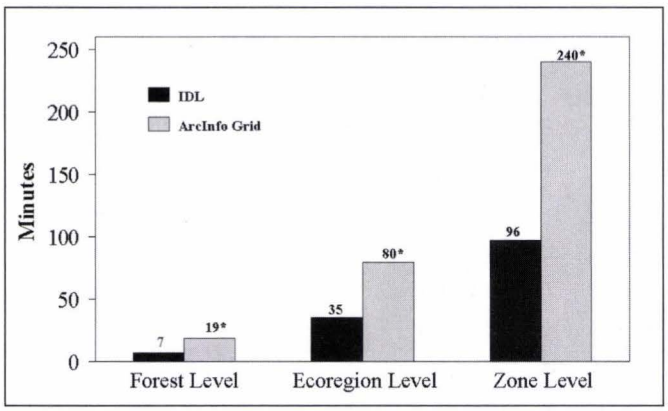
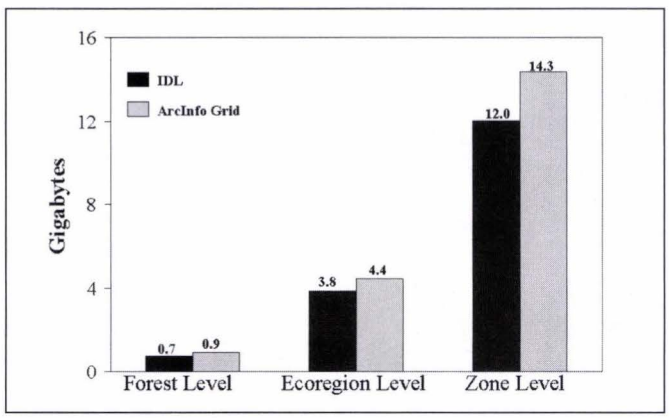


Figure 6.—*Comparison of required storage for the simple model in the ARCINFO and the IDL software packages.*



Comparison of ARCINFO and IDL Model Implementation Methods

The IDL method took about half as long as ARCINFO, although as the area of interest increased the difference lessened (figs. 4 and 5). The values for the IDL process included the time to convert the ARCINFO grids into binary files and the output binary file into an ARCINFO grid. The simple model was able to run in both ARCINFO and IDL, but the complex model resulted in erratic output in ARCINFO. Although binary files require more storage than ARCINFO grids, IDL input and output files required less total storage space than ARCINFO (fig. 6). The ARCINFO method created temporary intermediate grids that were saved to memory, whereas the IDL method processed one row of data at a time, thus requiring only the memory needed to hold one row.

Conclusion

All models were able to complete implementation in IDL and generate an output file. ARCINFO, while eventually able to generate an output file for the complex model, proved to be unreliable and unstable, often generating an output grid that showed all cell values as zero, or completely nonsensical data values. The feasibility of implementing IDL depended on more than technical results. Implementation required considering (1) the type of operating system required, (2) the computer hardware requirements, (3) the cost of obtaining and maintaining IDL, (4) the ease of use and access to people who know how to implement the software, and (5) the limited input data formats (ASCII or binary). We compared all of these aspects for ARCINFO and IDL (table 1). The advantages of ARCINFO are that many people know how to work with ARCINFO; model implementation is a simple, one-word command (&r <name of model AML>); and data inputs and output are grids. The disadvantages are that it can be unstable for complex models, and large amounts of memory are required because all input grids and intermediate grids are stored in memory. The advantages of IDL are that it is less expensive than ARCINFO; requires less memory because model implementation is done only on certain sections of the input files at a time, and the programming lan-

Table 1.—*Comparison of (1) operating systems that could run each package respectively, (2) machine requirements, (3) approximate cost, (4) potential number of experienced users, and (5) available input data formats for ARCINFO and IDL*

Comparison Factors	ARCINFO	IDL
Operating system	Windows (95, 98, 2000, NT)	Windows (95, 98, 2000, NT)
	Unix	Unix
Computer requirements		
RAM	16MB	256MB
Hard drive	25MB	1GB
Processor	486 or higher	Pentium II
Cost	$3,000 (Windows)	Windows $650/ Unix $1,250
Ease of use	Many routines written; many AML programmers	Some routines written; fewer IDL programmers
Data format	Grid, binary, ASCII	Binary, ASCII

guage is easy to learn. The disadvantages of IDL are that there are fewer people who know how to write IDL code than ARCINFO AML code, a program needs to be written to implement the model, and input and output data are restricted to binary or ASCII formats.

Literature Cited

Bailey, R.G. 1980. Description of the ecoregions of the United States. Misc. Publ. 1391. Washington, DC: Department of Agriculture.

Frescino, T.S.; Moisen, G.G. 2004. Predictive mapping of forest attributes on the Fishlake National Forest. In: Proceedings of the 4th Annual Forest Inventory and Analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Friedman, J.H. 1991. Multivariate adaptive regression splines. Annals of Statistics. 19: 1–141.

Using FIA and GIS Data to Estimate Areas and Volumes of Potential Stream Management Zones and Road Beautifying Buffers

Michal Zasada[1,5], Chris J. Cieszewski[2], Roger C. Lowe[3], Jarek Zawadzki[1,6], Mike Clutter[4], Jacek P. Siry[2]

Abstract.—Georgia Stream Management Zones (SMZ) are voluntary and have an unknown extent and impact. We use FIA data, Landsat TM imagery, and GAP and other GIS data to estimate the acreages and volumes of these buffers. We use stream data classified into trout, perennial, and intermittent, combined with DEM files containing elevation values, to assess buffers with widths consistent with Best Management Practice rules. Our results suggest that SMZs in Georgia occupy about 3.6 percent of the forested area and contain about 4 percent of its volume. Assuming 100-foot buffers, the area would be more than 7.5 percent and the volume 8.4 percent.

As the country becomes more populated, urban expansion will leave fewer acres available for production forestry. We will also face greater demand for clean water and other nontimber forest benefits, which will also reduce the number of acres available for production forestry. At the same time, demand for various wood products from our forests will increase. Thus, we will have less land from which more wood products will be required (Wear and Greis 2002). This could mean that the standing timber supply may not meet demand. Since policymakers and business leaders make decisions that affect our forests, we need tools to evaluate the potential effects that their decisions will have on this resource. For example, we might wish to know how environmental constraints (such as mandatory SMZs of various sizes along streams and rivers) would affect the inventory of merchantable trees or business decisions such as the location of pulp or sawmills in particular locations within the State.

Riparian / Stream Management Zones

The Stream Management Zone (SMZ) is a mandated protection zone around a stream, lake, or other water body meant to protect factors such as water quality and fish habitat. The SMZ consists mostly of riparian habitat area—the area directly adjacent to a waterway that includes the bank vegetation and a strip of forest.

The most important function of riparian zones is maintaining water quality. They buffer rivers from adjacent pollution sources by filtering sediments, absorbing nutrients, and stabilizing stream banks. Riparian zones also provide habitat for wetland animals and plants, and travel routes for others. They provide habitat and food for stream organisms, and, by shading streams, moderate ambient temperature (Welsch 1991).

Stream management zones were mandated by Federal water quality legislation ("Clean Water Act") to minimize or prevent nonpoint sources of water pollution (NPSP). In 1976 the U.S. Environmental Protection Agency recommended using Best Management Practices (BMPs) as a primary method for controlling NPSP. The State of Georgia chose a nonregulatory system of voluntary compliance. The Georgia Forestry Commission issued the current "Georgia's Best Management Practices For Forestry" manual in 1999.

Nowadays we are facing the possibility of mandatory BMPs for all forested areas. Changes in BMPs are also possible to meet the demands of some environmentalists to make required buffers around streams wider and forest management inside of them more restricted (Wenger 1999). Currently, for perennial streams BMPs recommend leaving evenly distributed 50 square feet of basal area per acre or at least 50 percent of the canopy cover after a harvest. For trout streams, an additional no-harvest zone around the stream's bank must be created. For intermittent streams, requirements include leaving 25 square feet of basal area per acre or at least 25 percent of

[1]Postdoctoral Fellow, [2]Assistant Professor, [3]GIS Analyst, [4]Professor, respectively, Warnell School of Forest Resources, University of Georgia, Athens, GA.
[5]Assistant Professor at Faculty of Forestry, Warsaw Agricultural University, Poland.
[6]Assistant Professor at the Environmental Engineering Department, Warsaw University of Technology, Poland.

canopy cover after a harvest (GFC 1999). The extent of these potential harvesting limitations is not yet known.

Although many studies on riparian / stream management zones in the South exist (e.g., Wenger 1999), literature on their extent, assessment, and statistics is scarce. Cubbage and Woodman (1993) estimated harvesting area losses and costs by management classes using hypothetical "Forest Management Units" (FMUs) based on data from the FIA inventory for Georgia. They estimated forested area in stream management zones to range from 4.8 percent of total forest area (based on recommendations from BMPs of 1989), through 5.3 percent with buffers 35 to 100 feet wide, depending on slope to 7.09 percent for primary SMZs 300 ft wide. Beyond this study, some limited statistics are available from small-scale assessments and experiments, including ecological, biochemical, physiological, socioeconomical, and hydrological issues (e.g., Coweeta Long Term Ecological Research). However, these studies use mostly simplified assumptions (area of the buffer is calculated as a product of the stream length and the buffer width) and don't provide more detailed statistics.

Specific Objectives

We analyzed various Georgia data by relating spatially the FIA data to Landsat TM imagery and other GIS data to estimate the acreages and volumes of the protective zones and to assess potential impacts of implementing these zones in Georgia. Our specific objectives were to:

1. Develop an extended spatially explicit database with the FIA inventory information, water resources, and road developments in Georgia, plus hypothetical information on areas potentially classified as SMZs and road buffers.
2. Use the above database to evaluate the effects of regulations on timber harvests at various distances from roads and water resources in the State.

Data

In our study we used several sources of data. Administrative boundaries, roads, rivers, and elevation data (Digital Elevation Model – DEM) were downloaded from the Georgia GIS Data Clearinghouse located at (www.gis.state.ga.us/Clearinghouse/clearinghouse.html). The Georgia GIS Data Clearinghouse provides access to Geographic Information Systems (GIS) resources of Georgia for use by government, academia, and the private sector, mostly free of charge.

Information about natural land cover types comes from Georgia GAP Analysis Program data and was obtained from the Internet (http://narsal.ecology.uga.edu/gap/gap_landcover.html). Coordinated under the National Gap Analysis Program of the USGS Biological Resources Division, GAP is a nationwide biological diversity assessment and planning program. It assesses the conservation status of native vertebrate species and natural land cover types throughout the U.S., and facilitates the application of this information to land management activities. The GAP data set distinguishes 18 general land cover types, including three forest cover types (deciduous, evergreen, and mixed), forested wetlands that can be also classified as deciduous forests, and clearcut/sparse areas, derived from 1998 Landsat TM imagery by a series of unsupervised classifications. In addition, several ancillary data sources were incorporated, including roads, power lines, and the National Wetlands Inventory data. Delineation of additional subclasses of vegetation is still in progress. We used this data set to stratify Landsat TM imagery (16 scenes of Georgia collected in 1999 and 2000) data into three forest types (evergreen, deciduous, and mixed) and other nonforest types, to mask out some water bodies and roads, and to determine the initial cover type inside stream and road buffers.

Trout streams were identified using a "Trout Streams of Georgia" data set from NARSAL, Institute of Ecology, College of Environment and Design.

Methods and Assumptions

First, data on streams were classified into trout, perennial, and intermittent. Trout streams are those suitable for trout (potentially carrying trout). Perennial streams flow in a well-defined channel throughout most of the year under normal weather con-

Table 1.—*Buffer widths according to Georgia's current BMPs*

Stream type	Slope	Buffer width
Trout	All	[ft]
		100
Perennial	Slight (<20%)	40
	Moderate (21-40%)	70
	Steep (>40%)	100
Intermittent	Slight (<20%)	20
	Moderate (21-40%)	35
	Steep (>40%)	50

ditions. Intermittent streams flow in a well-defined channel during wet seasons but not for the entire year. Classified streams were masked out with assumed widths. In addition to data from the Georgia GIS Clearinghouse, we added water bodies extracted from GAP (recognized on satellite images as water bodies).

DEM data sets contain elevation values that were converted to slopes expressed in percent. The slope data set was reclassified according to BMP recommendations as slight (0-20 percent), moderate (21-40 percent), and steep (>40 percent) slope using ArcView software. Having classified the streams and derived slopes, we could finally create buffers around streams using widths from the water body shown in table 1. Buffers were stored in the GIS software (ArcView and ArcInfo) as a separate information layer. Transportation class (roads) was extracted from GAP data and combined with GIS Clearinghouse data. We assumed primary and secondary buffers (40 and 100 feet) that were processed similarly to streams. Resulting buffered streams and roads were intersected with the GAP land cover information layer.

The next step included the use of GAP data to stratify Landsat imagery into broad cover type classes. The stratified Landsat data were classified using the Euclidian distance approach developed by the USDA Forest Service (Ruefenacht *et al.* 2002). Classification for each forest cover type class was performed separately. Resulting polygons got FIA data assigned and were intersected with buffered streams and roads. It allowed calculating more detailed statistics, than using first step based only on the GAP data set analysis.

Results

The first results are based on GAP data only. As a result, we obtained area characteristics by GAP cover types. We also combined areas with volume data. For each of the broad forest cover types we calculated average volume per area using the FIA database (Hansen *et al.* 1992, Miles *et al.* 2001) and multiplied it by the number of acres in a given class. Areas and volumes were compared with the latest results of the FIA program (Thompson 1998) and presented in table 2. Stream buffers established according to Georgia's BMPs occupy about 872,500 acres, which makes up 3.57 percent of total forested area of the State. Assuming all buffers to be 100 feet wide, we came up with over 7.5 percent of forested area. Forests in the determined buffers contain 3.96 and 8.39 percent of total inventoried volume, depending on buffer width. Results for the scenario based on BMPs are smaller than numbers coming from studies by Cubbage and Woodman (4.8 and 5.3 percent), although the authors of the cited study assumed different buffer widths and a completely different (aspatial) approach. The area of the buffers in "primary SMZ zones" (7.09 percent) corresponds to our results from the scenario assuming all buffers are 100 feet wide. Shares of various cover types from our study are comparable to corresponding management class shares from the cited study (see Cubbage and Woodman 1993).

The next step of the research will be done on classified Landsat TM data with spatially distributed FIA attributes. Due to their computational requirements, these analyses are unfinished.

Literature Cited

Cubbage, Frederick W.; Woodman, J.N. 1993. Impacts of streamside management zones on timber supply in Georgia. In: Proceedings, Southern forest economics workers meeting; 1993 April 23–24; Durham, NC: 130–137.

Georgia Forestry Commission. 1999. Georgia's best management practices for forestry. Georgia Forestry Commission.

Hansen, Mark H.; Frieswyk, T.; Glover, J.F.; Kelly, J.F. 1992. The eastwide forest inventory database: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 48 p.

Table 2.—*Summary statistics of stream buffers in the State of Georgia based on GAP data and FIA summaries*

GAP code	Code volume description	Stream buffers according to BMPs		Stream buffers 100 ft wide	
		Volume		Area	
		Area [acres (%)]	[mill.ft³ (%)]	[acres (%)]	[mill.ft³ (%)]
7	Beaches/dunes	918		2,139	
18	Transportation	38,291		83,332	
20	Utility swaths	3,448		7,736	
22	Low-intensity urban	16,433		37,775	
24	High-intensity urban	4,728		11,712	
31	**Clearcut/sparse**	**46,212 (0.19)**		**107,089 (0.44)**	
33	Quarries/strip mines	1,083		2,705	
35	Rock outcrop	22		42	
41	**Deciduous forest**	**301,191 (1.23)**	**539 (1.60)**	**576,466 (2.36)**	**1032 (3.07)**
42	**Evergreen forest**	**186,259 (0.76)**	**234 (0.70)**	**435,920 (1.79)**	**548 (1.63)**
43	**Mixed forests**	**84.157 (0.34)**	**106 (0.32)**	**156,078 (0.64)**	**197 (0.58)**
73	Golf courses	1,168		2,696	
80	Pasture	61,188		123,100	
83	Row crop	27,608		72,940	
91	**Forested wetland**	**253,915 (1.04)**	**455 (1.35)**	**584,716 (2.40)**	**1,046 (3.11)**
92	Coastal marsh	14,557		35,432	
93	Nonforested wetland	5,003		8,697	
	Total Forests	**871,734 (3.57)**	**1334 (3.96)**	**1,860,268 (7.62)**	**2,824 (8.39)**

Miles, Patrick D.; Brand, G.J.; Alerich, C.L.; *et al.* 2001. The forest inventory and analysis database: database description and user's manual, version 1.0, revision 8 (June 1, 2001). St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station (FIA program Web site).

Ruefenacht, B.; Vanderzanden, D.; Morrison, M.; Golden, M. 2002. New technique for segmenting images. U.S. Department of Agriculture, Forest Service, Remote Sensing Application Center, RSAC Program Manual. 14 p.

Thompson, Michael T. 1998. Forest statistics for Georgia, 1997. Resour. Bull. SRS-36. Asheville, NC: U.S. Department of Agriculture, Forest Service. Southern Research Station. 92 p.

Wear, David N.; Greis, J.G. 2002. Southern forest resource assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 635 p.

Welsch, David J. 1991. Riparian forest buffers; function and design for protection and enhancement of water resources. NA-PR-07-91. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Area State and Private Forestry.

Wenger, Seth J. 1999. A review of the scientific literature on riparian buffer width, extent and vegetation. Athens, GA: University of Georgia, Public Service & Outreach, Institute of Ecology. 59 p.

FOREST SERVICE
US
DEPARTMENT OF AGRICULTURE
JOIN US
1905-2005
100